Inversion acoustique tridimensionnelle des formes d'onde complètes : méthodes algorithmes et application au réservoir pétrolier de Valhall

Guanghui Hu

UNIVERSITÉ DE GRENOBLE

THÈSE

Pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité : **Géosciences, Terre Solide**

Arrêté ministériel : 07 août 2006

Présentée par

Guanghui HU

Thèse dirigée par **Jean Virieux**
et codirigée par **Stéphane Operto**

préparée au sein de l' **Institut des Sciences de la Terre**
et de **École doctorale Terre Univers Environnement**

Three-dimensional acoustic Full Waveform Inversion: method, algorithms and application to the Valhall petroleum field

21 Septembre 2012 ,
devant le jury composé de :

René-Édouard Plessix
Chercheur Senior à SHELL, Rapporteur

Hervé Chauris
Professeur à l'École des Mines de Paris, Rapporteur

Michel Dietrich
Directeur de Recherche au CNRS, ISTerre, Université Joseph Fourier, Examinateur

Matthias Delescluse
Maître de conférences, Géologie, ENS Ulm, Examinateur

Stéphane Operto
Chargé de Recherche au CNRS, GeoAzur, Université de Nice-Sophia Antipolis, Valbonne, Co-Directeur de thèse

Jean Virieux
Professeur, ISTerre, Université Joseph Fourier, Grenoble, Directeur de thèse

Remerciements

La réalisation de cette thèse fut une occasion merveilleuse de rencontrer et d'échanger avec de nombreuses personnes. Je tiens à remercier les différentes personnes, qui de près ou de loin, ont contribué à l'aboutissement de ce travail de thèse. Je m'excuse d'avance pour toute personne que j'aurais pu oublier...

Je dédie tout d'abord mes remerciements à mes directeurs de thèse Stéphane Operto et Jean Virieux, qui ont su me guider tout au long de ce long périple, pour avoir créé un environnement de travail dynamique et motivant, et pour m'avoir fait confiance pour mener à bien ce projet. Je tiens à les remercie pour leurs sympathies, leurs compréhensions, leurs suggestions, leurs conseils, leurs suivis, leurs implications et leurs supports continu tout au long de ma thèse.

Je tiens à remercier Hervé Chauris et René-Édouard Plessix d'avoir accepté d'être les rapporteurs de ce travail. Je remercie également Matthias Delescluse et Michel Dietrich d'avoir accepté de participer au jury.

Je souhaite particulièrement remercier Vincent Etienne avec qui nous avons suivi le travail de l'inversion 3D acoustique. Cette thèse n'aurait sûrement pas pu s'accomplir sans son aide.

Je tiens à remercier Romain Brossier pour l'ensemble de sa contribution. Merci de m'avoir assisté avec beaucoup de patience en début de thèse, et merci pour les nombreux emails échangés et nombreuses heures passé au téléphone pour résoudre mes problèmes.

Je remercie également tous les membres du group SEISCOPE au sein duquel j'ai effectué ma thèse. Ma sincère gratitude s'adresse à Alessandra Ribodetti pour sa gentillesse et ses conseils. Je remercie Yasser Gholami, Vincent Prieux, Damien Pageot, pour les fructueuses discussions. Je remercie Clara Castellanos avec qui j'ai partagé au bureau de nombreuses discussions autour de la méthode state-adjoint. Je remercie mes collègues à ISTerre: Amir Asnaashari, Bastien Dupuis, François Lavoué, Yuelian Jia, et Aurélien Roques, avec qui j'ai partagé les congrès, les *Annual Meeting* et les non moins fameuses *réunions SEISCOPE*.

Je tiens à remercier Alain Dujardin et Tram Ding Tronc avec qui j'ai partagé le bureau. Je les remercie pour partager les questions françaises, les inouvrables discussions. Je tiens également à remercier à mes copains chinoises à Grenoble: Jing He, Hanyu Li, Jing Hu. Je les remercie pour leurs amitiés, leurs avis éclairés et surtout leurs aides pendant que j'étais absent.

Je tiens à remercier les membres de ISTerre et Géoazur qui ont rendu mon séjour au sein du laboratoire très agréable. Je pense aux nombreux étudiants et chercheurs avec qui j'ai partagé les repas, les pauses café et des discussions scientifiques. Je remercie plus particulièrement le personnel informatique et administratif pour leur efficacité et leur gentillesse.

Enfin, je remercie ma famille: ma grand-mère Fenglan Sun, mes parents Yilian Ma et Chunxi Hu et mes soeurs pour leurs compréhensions, leurs consolations et leurs aides malgré

la distance qui nous sépare.

Et bien sûr, je remercie les sponsors de SEISCOPE sans qui cette thèse n'aurait pas été possible: BP, CGG Veritas, ENI, Exxon Mobil, BR Petrobras, Shell, Statoil, SaudiAramco et Total.

Résumé

L'imagerie quantitative des propriétés physiques du sous-sol est fondamentale pour de nombreuses applications impliquant des échelles d'exploration très variées: géotechnique pour l'imagerie de la proche surface, exploration à l'echelle crustale, reconstruction lithosphérique et imagerie globale pour la compréhension fondamentale des processus géodynamiques, mais aussi pour l'exploitation optimale des ressources du sous-sol.

Parmi les méthode géophysiques, les méthodes sismiques ont le pouvoir de résolution le plus élevé. La densification des dispositifs d'acquisition, la mise au point de sources et de capteurs large bande et l'augmentation de la puissance de calcul ouvrent de nouvelles perspectives pour le développement et l'application de méthodes non conventionnelles d'imagerie sismique pour une extraction plus complète de l'information contenue dans les données sismiques. Parmi ces méthodes d'imagerie non conventionnelles, les méthodes d'inversion du champ d'onde complet, fondées sur la résolution complète de l'équation d'onde pour le problème direct (modélisation sismique) et la résolution d'un processus d'optimisation pour le problème inverse, font actuellement l'objet de nombreux développements méthodologiques, tant au sein des communautés industrielles qu'académiques.

Le challenge numérique est la résolution du problème direct en trois dimensions pour un grand nombre de sources sismiques caractéristique des acquisitions pétrolières massives, et le challenge méthodologique est la gestion de la non-linéarité du problème inverse résultant de l'éclairage incomplet du sous-sol depuis la surface par des sources de bande-passante limitée. L'apport attendu de ces méthodes est la résolution de l'imagerie sismique de l'ordre de la demi-longueur d'onde propagée, sa capacité à imager des cibles complexes d'un point de vue structural notamment sous des écrans salifères ou basaltiques et la quantification des paramètres physiques caractérisant le sous-sol tels que la vitesse de propagation des ondes de compression à laquelle peuvent s'ajouter la densité, l'atténuation, la vitesse de propagation des ondes de cisaillement et des paramètres caractérisant l'anisotropie du milieu.

L'objectif de cette thèse est de poursuivre le développement d'une méthode d'imagerie sismique acoustique 3D par l'inversion du champ d'onde complet et de l'appliquer à des données réelles pétrolières 3D de fond de mer enregistrées sur le champ pétrolier de Valhall en Mer du Nord. L'inversion est effectuée en domaine fréquentiel où un nombre limité de fréquences est inversé suivant un protocole hiérarchique maintenant bien éprouvé procédant des basses fréquences vers les hautes fréquence: cette approche multi-échelle favorise la prise en compte de la non-linéarité du problème inverse.

L'approche de modélisation en domaine temporel avec extraction du champ monochromatique par une transformée de Fourier discrète est effectuée pour calculer les champs d'onde monochromatique nécessaires à la résolution du problème inverse. L'algorithme d'optimisation

du problème inverse est fondé sur une méthode de gradients conjugués préconditionés ou sur une méthode quasi-Newton. Les méthodes sont appliquées dans le cadre de l'approximation visco-acoustique isotrope où le milieu est paramétré par la vitesse de propagation des ondes de compression, l'atténuation et la densité. Seule, la composante hydrophone acquise en fond de mer est inversée. L'enjeu méthodologique de cette thèse est de fournir un modèle tridimensionelle du champ pétrolier de Valhall dans un cube de dimensions approximatives 18 km x 12 km x 5 km en poussant l'inversion à la fréquence la plus élevée possible.

L'objectif de cette application est de fournir une des premières évaluations du potentiel des méthodes d'inversion des formes d'onde pour l'imagerie de milieux géologiques 3D à partir de dispositifs d'acquisition modernes tels que ceux mis en œuvre par l'industrie pétrolière. Les moyens à mettre en oeuvre sont ainsi clarifiés par rapport aux démonstrations faites en recherche et développement industriel.

Abstract

Quantitative imaging of the subsurface physical properties is fundamental to many applications involving very various explorations, such as geotechnical imaging of the near surface, petroleum exploration, crustal lithospheric exploration. This helps us to understand the fundamental of geodynamic processes and also to exploit the resources of subsurface.

Among the geophysical methods, seismic methods can give a higher resolution. The improvements of the acquisition in size and density, the multifold/multicomponent wide-aperture and wide-azimuth acquisitions, and the increased high-performance computing power open new perspectives to develop and apply non-conventional seismic imaging methods for extraction more complete and continuous information in the seismic data. Among these non-conventional methods, the full waveform inversion method based on the complete resolution of the wave equation for the direct problem (seismic modeling) and the resolution of optimization process for the inverse problem, are currently the subject of many methodological developments, in both industrial and academic communities.

The numerical challenge is the resolution of the three-dimensional direct problem for a large number of seismic sources, typically few to tens of thousands in petroleum industry acquisition. The methodological challenge is the management of the non-linearity of the inverse problem resulting from the incomplete illumination of subsurface from the surface survey with a limited bandwidth source. The expected contribution of these methods is to reach a spatial resolution of half-a-wavelength. It has the ability to image complex structure targets such as saline or salt-bearing basaltic and to quantify the subsurface physical parameters such as velocity, density, attenuation, anisotropic parameters and so on.

The objective of this thesis is to develop a method of three-dimensional seismic imaging by full waveform inversion and apply it to real ocean-bottom data set recorded in the Valhall oil field (in the North Sea). The inversion is performed in frequency domain. A limited number of frequencies is inverted following a hierarchical protocol from low to high frequencies. This multi-scale approach helps to reduce the non-linearity of the inverse problem.

The modeling approaches is performed in time domain and monochromatic wavefields are extracted by discrete Fourier transform to solve the inverse problem in frequency domain. The optimization algorithm of the inverse problem is based on conjugate gradients method or quasi-Newton method. The method is applied in the framework of the visco-acoustic isotropic approximation, where the medium is parameterized by the velocity of compressional wave propagation, attenuation, and density. The hydrophone data component located at the seabed is inverted. The methodological issue of this thesis is to develop by full waveform inversion a three-dimensional high-resolution velocity model of the Valhall oil field in a cube with a size of 18 km $\times$ 12 km $\times$ 5 km, and to push the inversion towards frequencies as high as possible.

The purpose of this application is to provide an early evaluation of the potentialities of full waveform inversion for imaging three-dimensional geological environments from surface acquisition such as the ones carried out by the petroleum industry.

Contents

Introduction

Seismic imaging

The knowledge of the Earth as its internal compositions and structures should be considered at different scales. It is of major interest for economy, human livings, environmental purposes, and science. For example, the exploration of natural resources is currently a major economic issue for many countries. This exploration is more and more associated with optimal exploitation of these resources. Several geophysical methods and techniques have been developed for a quantitative estimation of these resources as various physical phenomena can hamper the interior of the Earth with different resolution. The main geophysical techniques currently in use are numerous and are based on electromagnetic fields, magnetic and electric static ones, gravimetric field, and seismic fields among others. The choice of one of these methods depends on the type of physical properties to be identified in the interior of the Earth and the related complexity of these structures. Optimal investigation depends also on the purpose of the search from economical reason to geodynamical understanding. In my thesis, I shall focus my attention on the seismic methods, which are known to have a high resolution.

Seismic exploration techniques are used to determine the geological and physical structures of the subsurface as a routine component in the search of hydrocarbon reservoirs. It is crucial to extract from the recorded data the physical parameters of the subsurface, typically the seismic wave propagation velocity, in order to correctly locate and estimate potential reservoirs. The seismic active acquisition experiment uses controlled sources as explosives, air guns or vibroseis trunks. These sources initiate waves inside the medium and these propagated waves are ultimately recorded at the surface by sensors at the receiver positions. They might be geophones, which record particle velocities along one up to three directions, or hydrophones which record the pressure component. During the propagation, the seismic wave undergoes at heterogeneities inside the medium several complex physical conversions as reflection, refraction and/or diffraction. These conversions, when recorded at the surface, bring information about the structure of the Earth one needs to interpret, and these conversions confer to seismic data non-linear properties making difficult if not impossible the reconstruction. Therefore, recorded signals should be processed in an adequate way for the model building representing geological structures of the subsurface.

The geometry of the seismic acquisition defines the target dimension one may expect: if the recording time is short, one can reconstruct only the shallow part of the Earth. If the seismic acquisition is deployed over a large distance requiring a longer time window of recording, one may extract information at deeper depths from the data. Indeed, wide aperture/azimuth and global offset acquisitions are necessary to record the diving waves in order to appropriately

image complex targets such as salt domes with dipping flanks for example. However, dense wide azimuth acquisitions have often presented financial and deployment challenges. In oil & gas exploration industry, dense multifold seismic reflection acquisition is the standard acquisition geometry especially in marine environment. The length of seismic streamer and cables has been increased from 2 km to more than 12 km and one ship might trace more than one streamer. This technological effort has been performed because many targets are under the sea water. Therefore, for a long time in reasonably complex structures, the data processing has been mainly based on reflected waves. The tracking of additional targets in more complex geological environments turns out to be difficult when reflections are highly deformed, because of faults, high-velocity variations, and so on.

Full waveform inversion - FWI

Quantitative seismic imaging of three-dimensional (3D) crustal structures is therefore one of main challenges of geophysical exploration at different scales for subsurface, oil exploration, crustal and lithospheric investigations. Full Waveform Inversion (FWI) is one of the most promising techniques for seismic imaging as acquisition improves in size and density. Since the pioneering work on full waveform inversion in 1980's (Tarantola, 1984a; Lailly, 1984), it has been developed both in the time and frequency domains. The frequency domain provides a natural framework to design multiscale imaging through successive inversions of increasing frequencies: proceeding sequentially from low to high frequencies defines a multi resolution imaging strategy, that helps to mitigate the non-linearity of the inverse problem (Pratt et Worthington, 1990; Pratt, 1999). Moreover, computationally efficient frequency domain full waveform inversion algorithms can be designed by limiting the inversion to a few discrete frequencies, when wide-aperture acquisition geometries are considered (Sirgue et Pratt, 2004).

Full waveform inversion is a challenging data-fitting procedure based on full wavefield modeling to extract quantitative information from seismograms. FWI was originally developed in the time domain (Tarantola, 1984a), whereas the frequency-domain approach was proposed mainly in the 1990s by G. Pratt and collaborators (Pratt, 1990a; Pratt et Worthington, 1990). The frequency-domain formulation of FWI has been shown to be effective to build accurate velocity models of complex structures from long-offset acquisition geometries (Ravaut *et al.*, 2004). The wide-azimuth acquisitions allow FWI to image the deeper parts of the medium using transmitted energy information from the data. All the information contained in the data is used to survey the subsurface physical properties beneath the zone of interest. As a result, FWI is a high resolution imaging process. It is able to provide a result with a spatial accuracy of half-a-wavelength (Sirgue et Pratt, 2004). FWI is based on a local optimization scheme, where the gradient of the misfit function can be computed efficiently with the adjoint-state method (Plessix, 2006; Castellanos *et al.*, 2011). However, FWI is an ill-posed problem, that requires the starting model to be close enough to the real one in order to converge to the global minimum. Another counterpart of FWI is the required computational resources when considering models and frequencies of interest. The task becomes even more challenging when one attempts to perform the inversion using the elastic equation (Shi *et al.*, 2007; Brossier *et al.*, 2009) instead of using the acoustic approximation (Mulder et Plessix, 2008; Barnes et Charara, 2009). In the last few years, due to the increase of the high performance computing power and some algorithmic enhancements, FWI has focused a lot of interests and continuous efforts towards inversion of 3D data sets at low frequencies. Remarkable applications have

been done in 3D using the acoustic approximation (Plessix, 2009; Sirgue *et al.*, 2010; Plessix et Perkins, 2010). However, further investigations are still required to understand which part of the wavefield is really exploited by acoustic FWI of wide-azimuth data in anisotropic environments. Velocity models built by FWI are conventionally used as background models for prestack depth migration (Ben Hadj Ali *et al.*, 2008). As such, the FWI velocity model should allow to flatten reflectors in common image gathers. However, with the development of wide-azimuth acquisitions, the ability of the FWI to exploit the full wavefield including diving waves and super-critical reflections deserves further quality control of the FWI results, in particular in anisotropic environments.

The choice between time and frequency domains in FWI

FWI can be implemented in the time or frequency domain for both the forward and inverse problem. As we noted, in 1980s, FWI was developed in time domain (Tarantola, 1984a). Gauthier *et al.* (1986) showed the first test examples with three different simple two-dimensional (2D) models, but its application to 3D real data has had to wait for almost two decades due to the computational cost. Then, in the following years, Mora (1987) and Crase *et al.* (1990) applied the time domain inversion on 2D elastic case. In time domain inversion, the data are represented by temporal seismograms. The early applications of this method suffered a prohibitive computational cost, which limits the possible iterations, and an inappropriate choice of short offset acquisition. These configurations limit the possibility of imaging the long and intermediate wavelengths and therefore make the process highly non linear. In the time domain, Bunks *et al.* (1995) proposed a multiscale FWI scheme, which can be more naturally implemented in the frequency domain. It consists of successive inversions of overlapping frequency groups. The first group contains only the starting frequency, and one higher frequency is added from one group to the next. This multiscale strategy successively inverts the subdata sets of increasing high-frequency content, because low frequencies are less sensitive to cycle-skipping artifacts and makes the problem more linear.

Frequency domain approach was proposed mainly in 1990s (Pratt, 1990a; Pratt et Worthington, 1990). The frequency domain FWI approach is equivalent to the time domain approach when all of the frequencies are inverted simultaneously (Pratt *et al.*, 1998). One of the most important advantages of frequency domain inversion is the ability to provide an unaliased image using a limited number of frequencies. The proposed strategy is very pertinent: the long offsets allow to rebuild the long wavelengths, which are indispensable for the convergence of the iterative system. A few discrete frequencies are selected for frequency domain FWI, and the inversion is carried out sequentially from low to high frequencies. It helps to reduce the non-linearity : the long wavelength components of the model parameters are recovered by low frequency, and more details and features are recovered as the inversion proceeds with higher frequencies. The starting model for the higher frequencies is the final recovered model by the previous frequencies. The second approach, which is referred to as the simultaneous inversion approach, consists of successive inversions of slightly overlapping frequency groups. The choice of the frequency bandwidth should consider the trade-off between computational efficiency and quality of imaging, the large bandwidth of the frequency can mitigate the non-linearity of FWI in terms of the non-unicity of the solution, whereas the maximum frequency of the group should be chosen by such that the cycle-skipping artifacts are avoided (Virieux et Operto, 2009). An example of this tuning is illustrated by Brossier *et al.* (2009). The frequency domain provides

a more natural framework for this multiscale approach by performing successive inversions of increasing frequencies (Virieux et Operto, 2009), while the strategy for forward modeling could be adapted to the available computational resources.

Non-linearity of FWI can also be efficiently mitigated by selecting a subset of specific arrivals (*i.e.*, early arrivals, reflected phases) in the data by time windowing (e.g., Sheng *et al.*, 2006; Sears *et al.*, 2008). Frequency domain wave modeling is not as flexible as the time-domain system for the preconditioning of the data by time windowing, as a limited number of frequencies is conventionally processed at a given step of the inversion. This makes frequency domain full waveform inversion based on time domain modeling an attractive strategy to design robust FWI algorithms.

Recently, due to the increase of the available computational power and some algorithmic enhancements, FWI has focused a lot of interests and continuous efforts towards inversion of 3D data sets at low frequencies. Remarkable applications have been done in 3D using the acoustic approximation (Plessix, 2009; Sirgue *et al.*, 2010; Plessix et Perkins, 2010). However, further investigations are still required to understand which part of the wavefield is really exploited by acoustic FWI of wide-azimuth data in anisotropic environments.

In the following of this thesis, we have chosen the strategy of inversion in frequency domain, while the forward problem will be performed in the time domain with an extraction of frequency response by discrete Fourier transform (Sirgue *et al.*, 2010). The FWI algorithm relies on a pseudo-conservative form of the velocity-stress wave equation. This allows first to make the kernel of the waveform inversion diagonal and independent of the numerical scheme used for seismic modeling, and hence to interface easily different modeling engines with the inversion. Second, the gradient of the misfit function can be built from self-adjoint operators, which allow us to use the same modeling scheme to compute the incident and adjoint wavefields. Two nested-levels of parallelism by source distribution and domain decomposition of the computational mesh allow us to optimize the performance of the code according to the computational platform.

Aim of the thesis

The objective of this thesis is to develop a method of 3D seismic imaging by frequency domain full waveform inversion based on time domain modeling and apply it to Valhall OBC real data set. The purpose of this application is to provide an early evaluation of the potentialities of full waveform inversion for imaging 3D geological environments from surface acquisition.

Summary of the thesis

This thesis will be organized in three parts. I will be concentrated in its first part on the forward problem as we need to solve it efficiently for many sources during the inversion procedure. We shall consider then the inversion problem formulation with the numerical difficulties we must face. Then, I shall consider the application on Valhall oil field.

Part I is devoted to forward problem. In the first chapter, I shall present the numerical solution of the wave equation in frequency and time domains. I aim to further understand the time and memory complexities for the two domains. Then I would like to validate our

time domain solution using $\mathcal{O}(\Delta x^4, \Delta t^2)$ finite-difference method with analytical solutions. I shall validate the implementation of the perfectly-matched-layer (PML) absorbing boundary condition, the free-surface boundary condition, and the source implementation on coarse grid with Sinc-function interpolations. Finally, I shall perform a scalability analysis of the domain-decomposition parallelism.

Part II is devoted to the inverse problem. In the second chapter, I am interested in the approach of frequency domain full waveform inversion based on time domain modeling. I shall present the line search methods used to solve a local optimization problem. I shall describe the method to build the gradient of the misfit function with the adjoint-state method. The least-squares formalism is reviewed. I shall discuss the two-nested levels of parallelism by source distribution and domain decomposition of the computational domain. The algorithm that combines the two level parallelisms, will be presented in this part. Finally I shall conclude this algorithm investigation with same validation tests.

Part III is devoted to application. Two applications of the algorithm to EAGE/SEG velocity model and the Valhall model will be presented in the third chapter. Before the application of Valhall real data set, the method is validated against the EAGE/SEG Overthrust model. I shall analyze our FWI model per frequency groups. I shall discuss the accuracy of initial model, the quality of data, the strategy of FWI, the choice of frequency group and so on.

In the last chapter, I shall expose the general conclusions and perspectives of the thesis.

Part I

Forward problem

In the first part of this thesis, I shall discuss the different approaches for solving efficiently partial differential equations in heterogeneous continuous media for seismic wave propagation. Seismic imaging needs quite efficient modeling tools for the forward problem in order to estimate the entire wavefield at each node of the computational domain considering all types of propagation (direct, reflected, diffracted, transmitted waves should be included in our simulation) in an heterogeneous medium. We have to perform this modeling for many sources, typically for thousands of them in three-dimensional setting. Moreover, in the framework of full waveform inversion, we have to repeat this procedure of many modelings at each iteration of the model update.

Firstly, I shall introduce the different equations handling the wave propagation and I shall focus my attention on volumetric methods in relation with partial differential equations. We consider these equations for elastic and acoustic wave propagation in order to appreciate differences between them. The main concern in this work is related to acoustic wave propagation.

Secondly, I concentrate my attention in solving second-order acoustic wave equation in the frequency domain directly in order to illustrate advantages and disadvantages for working in this specific frequency domain. I shall present the spatial stencils and the related sparse matrix I build. I shall discuss how to solve the related linear system efficiently. The memory requirement of the technique we solve leads us to consider another strategy (Brossier *et al.*, 2010a; Ben Hadj Ali *et al.*, 2008; Operto *et al.*, 2007; Sourbier *et al.*, 2011).

Thirdly, we investigate the time-domain formulation, which turns out to be less memory-demanding than the frequency approach available to us at the expense of computer times (Virieux *et al.*, 2009). I discuss the discrete implementation I consider as first-order acoustic wave equations. I validate the numerical solutions against analytical solutions in an infinite medium and in an half-space. I shall introduce in more details the absorbing boundary conditions (Komatitsch et Martin, 2007) as formulated in the time domain as well as the source implementation and the extraction of signals at receivers (Hicks, 2002). Finally, I shall analyze the capabilities for recovering monochromatic solutions through Discrete Fourier Transform (DFT).

Fourthly, I shall introduce two levels of parallelisms: one is over sources and it is related to embarrassing parallelism as we can start forward computing for each source independently and the other one is over domain decomposition as I may need more that one CPU for tackling the forward problem for one source. The second parallel implementation through a domain decomposition technique improves dramatically the efficiency of the forward modeling (Bohlen et Saenger, 2006), and it is crucial for seismic imaging.

Finally, I perform modeling in two realistic models. I consider first the EAGE/SEG Overthrust model and the EAGE/SEG Salt model. Comparisons between solutions at different frequencies between the second-order equation and the system of first-order equations will illustrate our capability of modeling accurately waves in complex media.

Chapter 1

Wave propagation in continuous medium

Contents

We shall consider the Earth as a mechanical continuous body where seismic wave propagation, as a physical phenomenon, obeys deterministic partial differential equations (**PDE**) we shall describe later on. Moreover, the Earth, and more specifically its superficial solid envelop named as the crust, is heterogeneous and dissipative regarding seismic waves. Luckily, the Earth has different heterogeneous scales as slow variations of velocities inside layers separated by discontinuities across which waves are partitioned. One may consider that different approaches could be considered from asymptotic methods as ray tracing methods quite valid in areas with smooth variations (Hanyga *et al.*, 1995; Hanyga et Pajchel, 1995; Hanyga et Seredyńska, 1999) to boundary element methods which discretize discontinuities only considering layers with homogeneous properties (Krishnasamy *et al.*, 1992; Aubry et Clouteau, 1991).

In seismic imaging, we could not have prior description about property variations and we, therefore, consider a grid approach with a volumetric description where both smooth and sharp variations should be sampled through grid nodes. These volumetric approaches rely on a 3D sampling of the medium, which could be on regular meshes as 2D or 3D rectangular grids or which could be on irregular grids based in 2D triangles/quadrilateral or 3D tetrahedral/hexahedral angles meshes. Triangle/tetrahedral meshes are the simplex unstructured grids one can think about. Regular grids are also very appreciated as they are easy to construct and as efficient approaches of solving PDE are often based on regular grids.

Full Waveform Inversion, which stands for a seismic imaging considering complete propagation features, requires accurate and efficient wavefield modeling engines, especially in 3D heterogeneous media. Finding complete (or full) accurate wavefield solution requires efficient numerical approaches and, because we are interested in heterogeneous media, we shall concentrate our attention to volumetric methods where the medium is discretized using a mesh approach. Efficient methods such as finite-differences approaches (Virieux, 1986; Levander, 1988; Graves, 1996; Operto *et al.*, 2007) may compete with more accurate methods such as pseudospectral approaches (Danecek et Seriani, 2008) or more elaborated methods such as continuous finite-elements approaches (Marfurt, 1984; Min *et al.*, 2003; Seriani et Priolo, 1994) or discontinuous finite-elements approaches (de la Puente *et al.*, 2008).

In this work, we shall mainly rely on simple forward modeling grids based on 3D rectangular prism (even cubic) grids on which partial differential equations need to be solved using finite-difference techniques as these methods are quite efficient especially for 3D geometries.

The modeling techniques are strongly dependent on the problem we need to consider and on the available computer resources. For seismic processing, finite-difference methods are widely used in the time and frequency domains, thanks to their efficiency and their simplicity. In an infinite continuous medium, the time-domain formulation relates inertial terms with internal forces and external forces leading to the following system of partial differential equations,

$$\mathbf{M}(\mathbf{x})\frac{\partial^2 \mathbf{u}(\mathbf{x},t)}{\partial t^2} = \mathbf{A}(\mathbf{x})\ \mathbf{u}(\mathbf{x},t) + \mathbf{s}(\mathbf{x},t), \tag{1.1}$$

where $\mathbf{M}$ and $\mathbf{A}$ are the mass and the stiffness matrices, respectively. The stiffness matrix exploits the Hooke law and expresses the linear differential operator between stresses and deformations linked to displacements. The source term is denoted by $\mathbf{s}(\mathbf{x},\mathbf{t})$, and the seismic wavefield is described by the quantity $\mathbf{u}(\mathbf{x},\mathbf{t})$, which represents the particle displacement vector. The system (1.1) is generally solved with an explicit time marching algorithm: the value of

the wavefield at time step (n+1) at a spatial position is inferred from values of the wavefield at previous time steps. If both velocity and stress wavefields are considered, the system of second-order equations can be recast into a first-order hyperbolic velocity-stress system as we shall see in the section 1.1.1.

We may take the Fourier transform $\mathbf{u}(\mathbf{x}, f)$ of the particle displacement $\mathbf{u}(\mathbf{x}, t)$, following the sign convention

$$\mathbf{u}(\mathbf{x}, f) = \int_{-\infty}^{\infty} \mathbf{u}(\mathbf{x}, t) e^{i\omega t} dt. \tag{1.2}$$

The Fourier convention is quite important to consider in the construction of the impedance matrix $\mathcal{B}$ from the matrices $\mathbf{M}$ and $\mathbf{A}$, especially when we shall consider the first-order system. In an infinite continuous medium, the frequency-domain formulation could be recast into the following linear system,

$$\mathcal{B}(\mathbf{x}, \omega)\mathbf{u}(\mathbf{x}, \omega) = \mathbf{s}(\mathbf{x}, \omega), \tag{1.3}$$

where $\mathcal{B}$ is the impedance matrix combining the mass and stiffness matrices while other terms have been already defined. The external source term $\mathbf{s}$ and the wavefield $\mathbf{u}$ are now expressed in the frequency domain.

The system of equations (1.3) can be solved by a decomposition of $\mathcal{B}$ such as lower and upper (LU) triangular decomposition, leading to direct-solver techniques. The advantage of the direct-solver approach is that, once the decomposition is performed, equation (1.3) is efficiently solved for multiple sources using forward and backward substitutions (Marfurt, 1984). The direct-solver approach is efficient for 2D forward problems. However, the time and memory complexities of LU factorization and its limited scalability on large-scale distributed memory platforms prevent use of the approach for large-scale 3D problems (*i.e.* problems involving more than 10 million unknowns (Operto *et al.*, 2007)), although recently Wang *et al.* (2011b) have introduced their study of the scalability and have proposed practical efficient parallel algorithms for direct solver based on hierarchically semiseparable (HSS) matrices. They showed a robust and efficient parallel implementation of the LU decomposition with promising perspectives.

In this work, we use an alternative approach proposed by Nihei et Li (2007) and Sirgue *et al.* (2008), who compute monochromatic wavefields in the time domain and extract the frequency-domain solution by discrete Fourier transform in the loop over the time steps. The advantages of this approach is that the memory complexity stays low but one has to perform the forward modeling for each source from scratch. During a forward modeling, an arbitrary number of frequencies can be extracted within the loop over time steps at minimal extra cost, allowing the grouping of selected frequencies for the inversion scheme.

Moreover, time windowing can be easily applied in the time domain, unlike in the frequency domain. Time windowing on both real and synthetic data allows the extraction of specific arrivals for FWI (early arrivals, reflections, post-critical reflections, converted waves), which is often useful to mitigate the non-linearity of the inversion by judicious data preconditioning (Brossier *et al.*, 2009; Fichtner *et al.*, 2008; Sears *et al.*, 2008).

In this section, I shall introduce the equations of wave propagation. These equations must be solved in arbitrarily heterogeneous media to constitute the direct problem, which is the engine of waveform inversion algorithms. I shall present the numerical method used to solve the elastodynamic equations in this work: finite-difference method for discretization of partial differential equations. I shall discuss the validations of this approach, and we will show the applications in realistic models: EAGE/SEG Overthrust model and EAGE/SEG Salt model.

1.1 The elastodynamic equations and the acoustic equations

In this section, I shall derive the well-known partial differential equations, which govern acoustic wave propagation for linear continuum media. It can be obtained through an acoustic approximation of the elastic system (Aki et Richards, 1980). In the elastic media, the wave equation is obtained by a combination of the equation of motion of Newtow and the Hooke law. The wave propagation is governed by the following linear elastodynamics equations, which ensure firstly the conservation of quantity of motion and, secondly, linear connection of strains and stresses in the material through the Hooke law.

1.1.1 First-order versus second-order formulations

The development of the equations of wave propagation can be found in many textbooks in physics (Duvaut, 1990; Royer et Dieulesaint, 1997) as well as in geophysics (Menke et Abbott, 1990; Shearer, 2009). The governing **PDE** system, called elastodynamic system, is written as the following first-order system,

$$\begin{aligned} \rho\partial_t \mathbf{v} &= \nabla\cdot\sigma + \mathbf{f}_{ext} \\ \partial_t \sigma &= \mathbf{c} : \nabla\mathbf{v}, \end{aligned} \tag{1.4}$$

with the velocity vector and stress tensor definitions

$$\begin{aligned} \mathbf{v} &= (v_x\, v_y\, v_z)^T \\ \sigma &= (\sigma_{xx}\, \sigma_{yy}\, \sigma_{zz}\, \sigma_{xy}\, \sigma_{xz}\, \sigma_{yz})^T. \end{aligned} \tag{1.5}$$

In the system (1.4), the external force is denoted by the term $\mathbf{f}_{ext}$, the density by ρ. The elastic tensor $\mathbf{c}$ contains more than 21 independent coefficients in the case of general (triclinic) anisotropic media. It is useful to detail these equations for isotropic elastic media as follows

$$\begin{aligned}
\frac{\partial\sigma_{xx}(x,y,z,t)}{\partial t} &= (\lambda(x,y,z)+2\mu(x,y,z))\frac{\partial v_x(x,y,z,t)}{\partial x} + \lambda(x,y,z)\{\frac{\partial v_y(x,y,z,t)}{\partial y} \\
&+ \frac{\partial v_z(x,y,z,t)}{\partial z}\} \\
\frac{\partial\sigma_{yy}(x,y,z,t)}{\partial t} &= (\lambda(x,y,z)+2\mu(x,y,z))\frac{\partial v_y(x,y,z,t)}{\partial y} + \lambda(x,y,z)\{\frac{\partial v_x(x,y,z,t)}{\partial x} \\
&+ \frac{\partial v_z(x,y,z,t)}{\partial z}\} \\
\frac{\partial\sigma_{zz}(x,y,z,t)}{\partial t} &= (\lambda(x,y,z)+2\mu(x,y,z))\frac{\partial v_z(x,y,z,t)}{\partial z} + \lambda(x,y,z)\{\frac{\partial v_x(x,y,z,t)}{\partial x} \\
&+ \frac{\partial v_y(x,y,z,t)}{\partial y}\} \\
\frac{\partial\sigma_{xy}(x,y,z,t)}{\partial t} &= \mu(x,y,z)\{\frac{\partial v_x(x,y,z,t)}{\partial y} + \frac{\partial v_y(x,y,z,t)}{\partial x}\} \\
\frac{\partial\sigma_{xz}(x,y,z,t)}{\partial t} &= \mu(x,y,z)\{\frac{\partial v_x(x,y,z,t)}{\partial z} + \frac{\partial v_z(x,y,z,t)}{\partial x}\} \\
\frac{\partial\sigma_{yz}(x,y,z,t)}{\partial t} &= \mu(x,y,z)\{\frac{\partial v_y(x,y,z,t)}{\partial z} + \frac{\partial v_z(x,y,z,t)}{\partial y}\}
\end{aligned}$$

$$
\begin{aligned}
\frac{\partial v_x(x,y,z,t)}{\partial t} &= \frac{1}{\rho(x,y,z)}\{\frac{\partial \sigma_{xx}(x,y,z,t)}{\partial x}+\frac{\partial \sigma_{xy}(x,y,z,t)}{\partial y}+\frac{\partial \sigma_{xz}(x,y,z,t)}{\partial z}\}\\
\frac{\partial v_y(x,y,z,t)}{\partial t} &= \frac{1}{\rho(x,y,z)}\{\frac{\partial \sigma_{xy}(x,y,z,t)}{\partial x}+\frac{\partial \sigma_{yy}(x,y,z,t)}{\partial y}+\frac{\partial \sigma_{yz}(x,y,z,t)}{\partial z}\}\\
\frac{\partial v_z(x,y,z,t)}{\partial t} &= \frac{1}{\rho(x,y,z)}\{\frac{\partial \sigma_{xz}(x,y,z,t)}{\partial x}+\frac{\partial \sigma_{yz}(x,y,z,t)}{\partial y}+\frac{\partial \sigma_{zz}(x,y,z,t)}{\partial z}\}, \quad (1.6)
\end{aligned}
$$

where λ and μ denote the Lamé coefficients. We may express this system in a matrix form: the velocity and stress components will be estimated at different time steps in order to allow time discretization using an integration scheme of leapfrog type. We have the compact expression

$$
\begin{aligned}
\partial_t \mathbf{v} &= \sum_{\theta\in\{x,y,z\}} \partial_\theta (A_\theta \sigma)\\
\partial_t \sigma &= \sum_{\theta\in\{x,y,z\}} \partial_\theta (B_\theta \mathbf{v}), \quad (1.7)
\end{aligned}
$$

with

$$
\begin{aligned}
A_x &= \begin{pmatrix} \frac{1}{\rho} & 0 & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & \frac{1}{\rho} & 0 & 0\\ 0 & 0 & 0 & 0 & \frac{1}{\rho} & 0 \end{pmatrix} \qquad B_x = \begin{pmatrix} \lambda+2\mu & \lambda & \lambda & 0 & 0 & 0\\ 0 & 0 & 0 & \mu & 0 & 0\\ 0 & 0 & 0 & 0 & \mu & 0 \end{pmatrix}^T\\
A_y &= \begin{pmatrix} 0 & 0 & 0 & \frac{1}{\rho} & 0 & 0\\ 0 & \frac{1}{\rho} & 0 & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & \frac{1}{\rho} \end{pmatrix} \qquad B_y = \begin{pmatrix} 0 & 0 & 0 & \mu & 0 & 0\\ \lambda & \lambda+2\mu & \lambda & 0 & 0 & 0\\ 0 & 0 & 0 & 0 & 0 & \mu \end{pmatrix}^T\\
A_z &= \begin{pmatrix} 0 & 0 & 0 & 0 & \frac{1}{\rho} & 0\\ 0 & 0 & 0 & 0 & 0 & \frac{1}{\rho}\\ 0 & 0 & \frac{1}{\rho} & 0 & 0 & 0 \end{pmatrix} \qquad B_z = \begin{pmatrix} 0 & 0 & 0 & 0 & \mu & 0\\ 0 & 0 & 0 & 0 & 0 & \mu\\ \lambda & \lambda & \lambda+2\mu & 0 & 0 & 0 \end{pmatrix}^T.
\end{aligned}
$$

As we are interested in the acoustic wave equation, we consider that there is no shear motion and, therefore, the second Lamé coefficient, known also as the shear modulus, is set to zero everywhere. Consequently, only normal components of stresses are different from zero values. The system (1.6) reduces to the following system

$$
\begin{aligned}
\frac{\partial \sigma_{xx}(x,y,z,t)}{\partial t} &= \lambda(x,y,z)\{\frac{\partial v_x(x,y,z,t)}{\partial x}+\frac{\partial v_y(x,y,z,t)}{\partial y}+\frac{\partial v_z(x,y,z,t)}{\partial z}\}\\
\frac{\partial \sigma_{yy}(x,y,z,t)}{\partial t} &= \lambda(x,y,z)\{\frac{\partial v_y(x,y,z,t)}{\partial y}+\frac{\partial v_x(x,y,z,t)}{\partial x}+\frac{\partial v_z(x,y,z,t)}{\partial z}\}\\
\frac{\partial \sigma_{zz}(x,y,z,t)}{\partial t} &= \lambda(x,y,z)\{\frac{\partial v_z(x,y,z,t)}{\partial z}+\frac{\partial v_x(x,y,z,t)}{\partial x}+\frac{\partial v_y(x,y,z,t)}{\partial y}\}\\
\rho(x,y,z)\frac{\partial v_x(x,y,z,t)}{\partial t} &= \frac{\partial \sigma_{xx}(x,y,z,t)}{\partial x}\\
\rho(x,y,z)\frac{\partial v_y(x,y,z,t)}{\partial t} &= \frac{\partial \sigma_{yy}(x,y,z,t)}{\partial y}\\
\rho(x,y,z)\frac{\partial v_z(x,y,z,t)}{\partial t} &= \frac{\partial \sigma_{zz}(x,y,z,t)}{\partial z}. \quad (1.8)
\end{aligned}
$$

We find that the first three equations of system (1.8) are equal, which $\partial\sigma_{xx}(x,y,z,t)/\partial t = \partial\sigma_{yy}(x,y,z,t)/\partial t = \partial\sigma_{zz}(x,y,z,t)/\partial t$. Therefore, let us define the pressure field $P(x,y,z,t)$

by the following expression

$$p(x,y,z,t) = \frac{\sigma_{xx}(x,y,z,t) + \sigma_{yy}(x,y,z,t) + \sigma_{zz}(x,y,z,t)}{3}.$$

Combining the system (1.8) and (1.9), we can obtain the first-order hyperbolic acoustic system as below:

$$\begin{aligned}
\frac{\partial p(x,y,z,t)}{\partial t} &= \kappa(x,y,z)\left(\frac{\partial v_x(x,y,z,t)}{\partial x} + \frac{\partial v_y(x,y,z,t)}{\partial y} + \frac{\partial v_z(x,y,z,t)}{\partial z}\right) \\
&+ s(x,y,z,t) \\
\frac{\partial v_x(x,y,z,t)}{\partial t} &= b(x,y,z)\frac{\partial p(x,y,z,t)}{\partial x} \\
\frac{\partial v_y(x,y,z,t)}{\partial t} &= b(x,y,z)\frac{\partial p(x,y,z,t)}{\partial y} \\
\frac{\partial v_z(x,y,z,t)}{\partial t} &= b(x,y,z)\frac{\partial p(x,y,z,t)}{\partial z},
\end{aligned} \tag{1.9}$$

where the quantity $\kappa(x,y,z)$ denotes the bulk modulus and the buoyancy $b(x,y,z) = 1/\rho(x,y,z)$ is the inverse of the density. The source term $s(x,y,z,t)$ is applied in the pressure field as we often consider explosive sources. This system is often called the system of vectorial acoustic wave equations. Please note that material properties could be easily move to the left-hand side of the system, separating them from spatial derivatives of the different fields: this is the so-called pseudo-conservative form which comes naturally for the acoustic system.

In the frequency domain, the system (1.9) can be transformed into

$$\begin{aligned}
p(x,y,z,\omega) &= \frac{\kappa(x,y,z)}{-\iota\omega}\left(\frac{\partial v_x(x,y,z,\omega)}{\partial x} + \frac{\partial v_y(x,y,z,\omega)}{\partial y} + \frac{\partial v_z(x,y,z,\omega)}{\partial z}\right) \\
&+ s(x,y,z,\omega) \\
\frac{-\iota\omega}{b(x,y,z)} v_x(x,y,z,\omega) &= \frac{\partial p(x,y,z,\omega)}{\partial x} \\
\frac{-\iota\omega}{b(x,y,z)} v_y(x,y,z,\omega) &= \frac{\partial p(x,y,z,\omega)}{\partial y} \\
\frac{-\iota\omega}{b(x,y,z)} v_z(x,y,z,\omega) &= \frac{\partial p(x,y,z,\omega)}{\partial z},
\end{aligned} \tag{1.10}$$

where $\iota = \sqrt{-1}$. This system allows the computation of the pressure and the particle velocity fields in the frequency domain. Please, note that the Fourier convention is quite important.

The second-order elliptical wave equation is obtained simply by injecting the last three equations into the first one of the system (1.9). It leads to

$$\left[\frac{1}{k(x,y,z)}\frac{\partial^2}{\partial t^2} + \frac{\partial}{\partial x}b(x,y,z)\frac{\partial}{\partial x} + \frac{\partial}{\partial y}b(x,y,z)\frac{\partial}{\partial y} + \frac{\partial}{\partial z}b(x,y,z)\frac{\partial}{\partial z}\right] p(x,y,z,t) = s(x,y,z,t). \tag{1.11}$$

In the frequency domain, the differential equation (1.11) becomes

$$\left[\frac{\omega^2}{\kappa(x,y,z)} + \frac{\partial}{\partial x}b(x,y,z)\frac{\partial}{\partial x} + \frac{\partial}{\partial y}b(x,y,z)\frac{\partial}{\partial y} + \frac{\partial}{\partial z}b(x,y,z)\frac{\partial}{\partial z}\right] p(x,y,z,\omega) = s(x,y,z,\omega). \tag{1.12}$$

The equation can be expressed in a compact form as

$$\mathcal{B}(x,y,z,\omega)p(x,y,z,\omega) = s(x,y,z,\omega), \tag{1.13}$$

where $\mathcal{B}$ denotes the spatial differential operator we need to discretize for the construction of a linear system to be solved. If density is constant, the system (1.11) can be simplified to

$$\frac{\partial^2 p(x,y,z,t)}{\partial t^2} + v^2(x,y,z)(\frac{\partial^2 p(x,y,z,t)}{\partial x^2} + \frac{\partial^2 p(x,y,z,t)}{\partial y^2} + \frac{\partial^2 p(x,y,z,t)}{\partial z^2}) \quad = \quad s(x,y,z,t), \tag{1.14}$$

where $v(x,y,z)$ is the wave speed. We have assumed that the density, and therefore the buoyancy, is constant. This equation is often called the acoustic wave equation.

1.1.2 Pseudo-conservative form of the elastodynamic equations

When performing seismic imaging, we have to consider not only the discretization of the forward problem but also the discretization of the adjoint problem as we shall see later. Some partial differential operators are self-adjoint, reducing our task for the numerical implementation. The first-order differential system is not self-adjoint for the elastic case. This is the reason why we have put our attention to an alternative form of the elastodynamics, namely, the pseudo-conservative form, which will be used for solving the adjoint problem for an isotropic medium using the discretization of the forward problem without a specific discretization of the adjoint problem (Castellanos *et al.*, 2011).

Let us consider the system equations of (1.7): we apply a transformation of the combined velocity-stress vector $\mathbf{u}$ into a new field $\mathbf{w}$ such that

$$\mathbf{w} = \mathbf{T}\ \mathbf{u}, \tag{1.15}$$

where the transformation matrix $\mathbf{T}$ is given by

$$\mathbf{T} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & 0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{1.16}$$

The vector u is noted by

$$u^t = (v_x, v_y, v_z, \sigma_{xx}, \sigma_{yy}, \sigma_{zz}, \sigma_{xy}, \sigma_{xz}, \sigma_{yz})^t. \tag{1.17}$$

Therefore, the new vector w can be written as

$$w^t = (v_x, v_y, v_z, \frac{1}{\sqrt{3}}\mathrm{Tr}(\sigma), \frac{\sqrt{3}}{\sqrt{2}}(\sigma_{zz} - \frac{1}{3}\mathrm{Tr}(\sigma)), \quad \tfrac{1}{\sqrt{2}}(-\sigma_{xx} + \sigma_{yy}), \sqrt{2}\sigma_{xy}, \sqrt{2}\sigma_{xz}, \sqrt{2}\sigma_{yz})^t.$$

(1.18)

The velocity components are not changed as well as shear components of the stress tensor. Let us introduce a matrix $\mathbf{P}$

$$\mathbf{P} = \begin{pmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & 0 & 0 & 0 \\ -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & 0 & 0 & 0 \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

The matrix $\mathbf{T}$ can be written as

$$\mathbf{T} = \begin{pmatrix} \prod_{3\times 3} & 0 \\ 0 & P \end{pmatrix}. \tag{1.19}$$

Multiplying an identity matrix noted as $P^{-1}P$ to the σ component in the system (1.7) and using $\tilde{\sigma} = P\sigma$, we have the following system

$$\begin{aligned} \partial_t v &= \sum_{\theta=\{x,y,z\}} \partial_\theta A_\theta P^{-1} \tilde{\sigma} \\ \partial_t \tilde{\sigma} &= \sum_{\theta=\{x,y,z\}} \partial_\theta P B_\theta v. \end{aligned} \tag{1.20}$$

We can now define the matrices Λ_1 and Λ_2 which contain the physical parameters as

$$\begin{aligned} \Lambda_1 &= diag\,(\rho, \rho, \rho) \\ \Lambda_2 &= diag\left(\frac{1}{3\lambda + 2\mu}, \frac{1}{2\mu}, \frac{1}{2\mu}, \frac{1}{\mu}, \frac{1}{\mu}, \frac{1}{\mu}\right). \end{aligned} \tag{1.21}$$

The system (1.20) can be written as

$$\begin{aligned} \Lambda_1 \partial_t v &= \sum \partial_\theta A'_\theta \tilde{\sigma} \\ \partial_t \tilde{\sigma} &= \Lambda_2^{-1} \sum \partial_\theta B'_\theta v, \end{aligned} \tag{1.22}$$

where operators A'_θ and B'_θ no longer contain physical parameters of the medium. After transferring the physical parameters on the left side, we have the pseudo-conservative system form

$$\begin{aligned} \Lambda_1 \partial_t v &= \sum_{\theta=\{x,y,z\}} A'' \tilde{\sigma} \\ \Lambda_2 \partial_t \tilde{\sigma} &= \sum_{\theta=\{x,y,z\}} B'' v, \end{aligned}$$

where $B'' = \sum \partial_\theta B'_\theta$ is explicitly defined as

$$B'' = \begin{pmatrix} \frac{1}{\sqrt{3}}\partial_x & \frac{1}{\sqrt{3}}\partial_y & \frac{1}{\sqrt{3}}\partial_z \\ -\frac{1}{\sqrt{6}}\partial_x & -\frac{1}{\sqrt{6}}\partial_y & \frac{\sqrt{2}}{\sqrt{3}}\partial_z \\ -\frac{1}{\sqrt{2}}\partial_x & \frac{1}{\sqrt{2}}\partial_y & 0 \\ \partial_y & \partial_x & 0 \\ \partial_z & 0 & \partial_x \\ 0 & \partial_z & \partial_y \end{pmatrix},$$

and $A'' = \sum \partial_\theta A'_\theta = B''^T$. The pseudo-conservative form can be written in a condensed form as

$$\Lambda \partial_t \mathbf{w} = \mathcal{B}' \mathbf{w}. \tag{1.23}$$

In the frequency domain, this system becomes

$$-i\omega \Lambda \mathbf{w} = \mathcal{B}' \mathbf{w}, \tag{1.24}$$

where matrices are defined following the different previous expressions as

$$\Lambda = \begin{pmatrix} \Lambda_1 & 0_{3\times 3} \\ 0_{6\times 3} & \Lambda_2 \end{pmatrix}, \quad \mathcal{B}' = \begin{pmatrix} 0_{3\times 3} & \mathbf{A}'' \\ \mathbf{B}'' & 0_{6\times 6} \end{pmatrix}, \quad \mathcal{B}' = \Lambda \mathbf{T} \mathcal{B} \mathbf{T}^{-\mathbf{1}}.$$

We have obtained a system where ${\mathcal{B}'}^t = \mathcal{B}'$ and ${\mathcal{B}'}^\dagger = -\mathcal{B}'$, thanks to the conservative property of the symmetrical matrix $\mathcal{B}'$. The appearance of the signal minus means that the problem has to be solved from the final time to the initial time. In the frequency approach, we must consider also this sign which leads to a 180^0 phase shift, an effect as important as the convention of sign for the Fourier transform. Therefore, we must consider the adjoint field of the field $\mathbf{w}$ which follows exactly the same PDE because the system is self-adjoint. Instead of doing so, we shall focus on the more familiar system for getting both the solution $\mathbf{u}$ and its adjoint, thanks to the transformation $\mathbf{T}$. Of note, this transformation is exactly the identity for the acoustic case. We will discuss further this point in the chapter on the inversion 2.1.3.

This discussion is valid for the elastic wave propagation and, consequently, could be considered as well for the acoustic one as we are going to solve the first-order hyperbolic system. We now proceed first in the discretization of the frequency approach of the acoustic second-order equation using a mixed-grid approach and, then, in the discretization of the time approach using a staggered-grid approach of the pressure/velocity field $\mathbf{u}$.

1.2 Frequency-domain methods

In the frequency-domain, the spatial discretization of partial differential equations reduces to the resolution of a complex-valued large and sparse system of linear equations for each frequency. The solution is the monochromatic wavefield when we consider the source excitation at this frequency as the right-hand side (r.h.s) of the linear system to be solved. The linear matrix system can be written as $\mathcal{B}\mathbf{p} = \mathbf{s}$. Two key issues should be addressed in frequency-domain wave modeling based on LU factorization of the impedance matrix: the first one is the memory required for the factorization of the matrix $\mathcal{B}$, while the second one is the poor scalability of such factorization over a cluster.

Although this work is not concerned directly by the frequency forward modeling, we shall compare our own monochromatic solution built from a time-domain formulation to the one directly computed by solving the frequency-domain system. Therefore, we shall describe shortly how the finite-difference (**FD**) stencil has been designed for efficient three-dimensional computation when considering only a frequency-domain approach.

In the frequency-domain, several approaches are available to solve the linear matrix system: the direct solver method (**DSM**) through an *LU* decomposition (Press *et al.*, 1992; Operto *et al.*, 2007), the iterative solver method (**ISM**) (Riyanti *et al.*, 2007; Plessix, 2007), and the hybrid solver method (**HSM**) based on domain decomposition (Haidar, 2008; Sourbier *et al.*, 2008) we shall discuss in this paragraph.

Direct solver methods are methods of choice when we can afford the memory requirement, because this approach is known for its multiple r.h.s resolution efficiency, an essential feature in the prospect of imaging where a large number of seismic sources is involved. A sparse direct solver performs first a lower-upper (LU) decomposition of the matrix, which is independent of the source, followed by forward and backward substitutions for each source in order to get the solution (Duff *et al.*, 1986).

Another approach for frequency-domain modeling is based on an iterative solver, the main advantage of this approach with respect to **DSM** is the small memory requirement, typically $\mathcal{O}(N^3)$ for 3D. Nevertheless, the performance of iterative methods depends strongly on the spectral properties of the linear system to be solved. In order to improve efficacy and robustness, an efficient preconditioning needs to be found. This task is critical and can be quite cumbersome (Plessix, 2009).

The third class of solvers, hybrid methods, may provide a good compromise between **DSM** and **ISM** in terms of memory requirement and efficiency of multiple r.h.s simulation (Virieux *et al.*, 2009). It tries to find a compromise between the two previous solvers through a domain decomposition method (Sourbier *et al.*, 2011).

We shall consider in this short description only the **DSM** approach as we have used these numerical solutions for comparison.

1.2.1 Spatial discretization

In **FD** methods, high-order accurate stencils are generally designed to achieve the best trade-off between accuracy and computational efficiency (Dablain, 1986). However, the **DSM** method needs to use high-order accurate stencils but their large spatial support will lead to a prohibitive

fill-in of the matrix during the LU decomposition (Hustedt *et al.*, 2004). Therefore, we need to preserve the compactness of stencils. The mixed-grid method has been proposed by Jo *et al.* (1996) to design both accurate and compact **FD** stencils. They have combined the standard Cartesian stencil and the 45 ^{0} rotated stencil (Saenger *et al.*, 2000) for a 2D geometry. The strategy is implemented with a parsimonious staggered grid approach by Hustedt *et al.* (2004) for acoustic wave equation and Stekl et Pratt (1998) developed similar discretisation for the elastic heterogeneous wave equation. In 3D geometry, three coordinate systems have been identified leading a a compact discrete operator with 27 coefficients (Operto *et al.*, 2007).

1. One standard Cartesian coordinate system, denoted $\mathcal{S}_{\mathcal{S}}$ (Figure 1.1-a).
2. Three coordinate systems, each one obtained by a 45° rotation of one of the axes of the standard coordinates system, denoted $\mathcal{S}_{\mathcal{R}}$ (Figure 1.1-b).
3. Four coordinate systems, each one obtained by considering only three axes from the four big cube diagonals, denoted $\mathcal{S}_{\mathcal{D}}$ (Figure 1.1-c).

The different stencils are mixed such that

$$w_1\mathcal{S}_{\mathcal{S}} + w_2\mathcal{S}_{\mathcal{R}} + w_3\mathcal{S}_{\mathcal{D}} = s, \tag{1.25}$$

where the weights w_1, w_2 and w_3 should verify the relationship

$$w_1 + \frac{w_2}{3} + \frac{w_3}{4} = 1. \tag{1.26}$$

The pattern of the impedance matrix inferred from the 3D mixed-grid stencil is shown in figure 1.2. The bandwidth of the matrix is of the order N^2 (N denotes one dimension of a 3D cubic N^3 domain). We keep it minimal thanks to the use of accurate compact stencils.

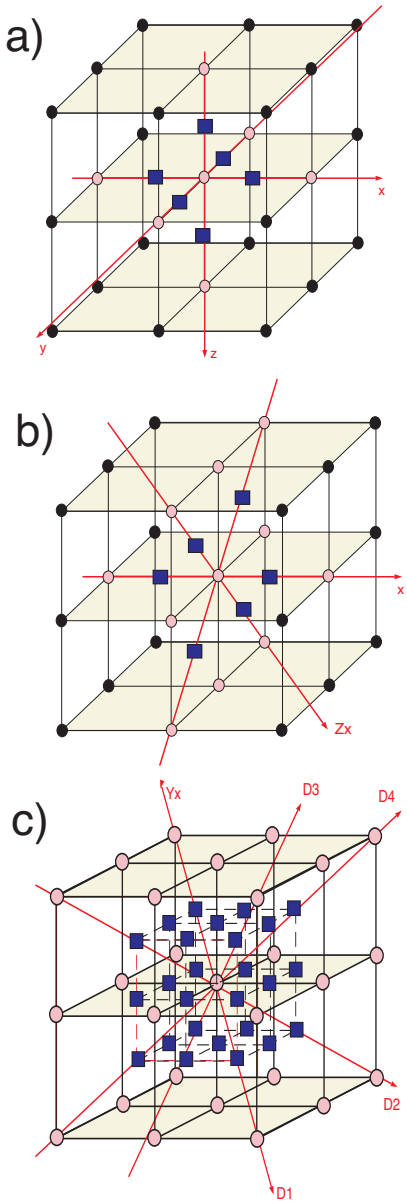

Figure 1.1: Different 3D finite difference stencils involved in the mixed-grid strategy. Circles are pressure grid points. Squares are positions where buoyancy needs to be interpolated in virtue of the staggered grid geometry. Pink circles are pressure grid points involved in the stencil. (a) Stencil on the classic Cartesian coordinate system. This stencil incorporates 7 coefficients. (b) Stencil on the rotated Cartesian coordinate system. Rotation is applied around the axis x in the figure. This stencil incorporates 11 coefficients. Same strategy can be applied by rotation around y and z. Averaging of the 3 resultant stencils defines a 19-coefficient stencil. (c) Stencil obtained from 4 coordinate systems, each of them being associated with 3 main diagonals of a cubic cell. This stencil incorporates 27 coefficients (Operto *et al.*, 2007)

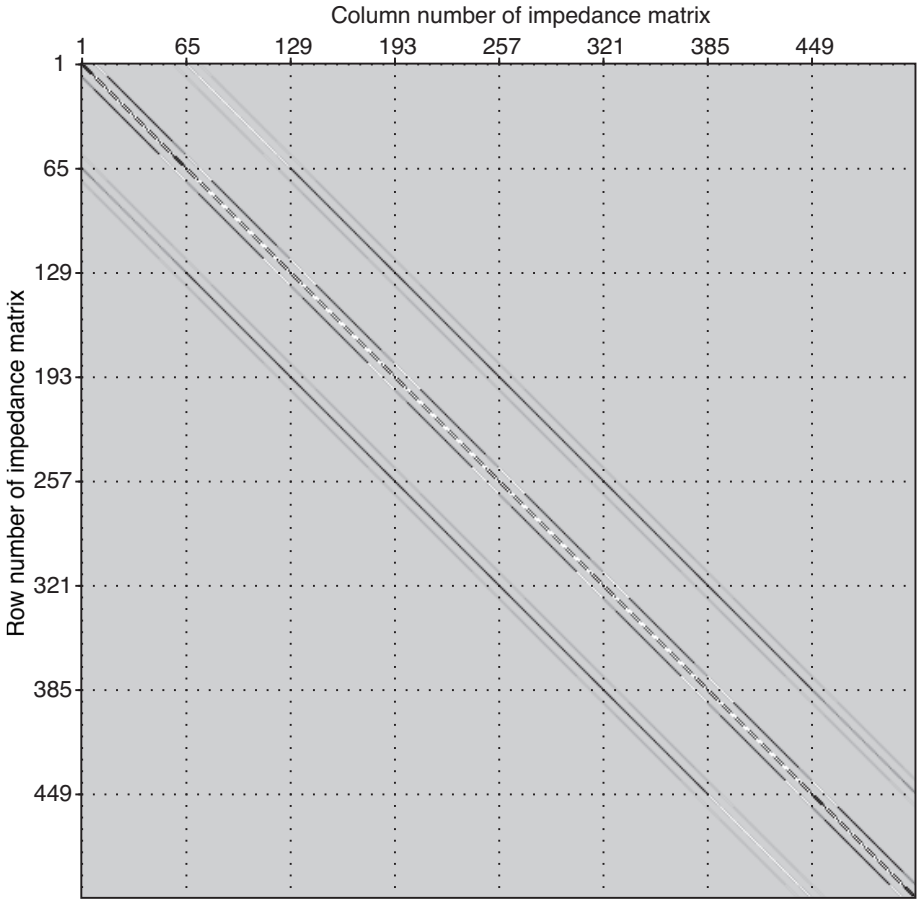

Figure 1.2: 3-D finite difference matrix, with 27 non-zero terms per row (Operto *et al.*, 2007). The matrix is band-diagonal with fringes. The bandwidth is $\mathcal{O}(2N_1N_2)$ where N1 and N2 are the two smallest dimensions of the 3D grid. The number of rows/columns in the matrix is $n_x \times n_y \times n_z = 8 \times 8 \times 8$.

1.2.2 Direct-solver approach

The direct solver method is the most accurate and quite robust approach for solving a linear system as long as the factorization can be realized. The solution precision is up to now the machine precision. The main advantage of this approach is the unique precomputation of the factorization at a given frequency for a large number of sources. Nevertheless, this approach is nowadays limited to sparse matrices problems of a few millions of unknowns due to its numerical cost in terms of CPU time and memory storage requirements and to limitations related to matrix conditioning issues. Unfortunately, the condition number increases when the size of the matrix and the associated physical problem increases. Recent investigations (Wang *et al.*, 2011a, 2010) may open doors for new perspectives as we may not need the machine precision accuracy.

The **DSM** methods are based on Gauss elimination technique. The main ideal of those methods is transforming the system $\mathcal{B}\mathbf{x} = \mathbf{s}$ into $(\mathbf{LU})\mathbf{x} = \mathbf{s}$. The matrix $\mathbf{L}$ is a lower triangular matrix and the matrix $\mathbf{U}$ is an upper triangular matrix for an unsymmetrical matrix. This system is then efficiently solved in two steps, forward and backward elimination phases, through inserting the temporary vector $\mathbf{y}$, $\mathbf{Ly} = \mathbf{s}$, and then $\mathbf{Ux} = \mathbf{y}$.

For sparse matrices, only non-zero matrix terms are stored. In the same way, only non zero-terms introduced in the LU decomposition are computed. However, the matrix decomposition leads to L and U matrices denser than the initial matrix and less than the full one. This issue is called the fill-in of the matrix. During the last decades, many techniques to reduce the fill-in have been developed. These techniques renumber/reorder the rows/columns of the matrix based on its graph. For these reasons, these techniques are called reordering techniques (George et Liu, 1981; Amestoy *et al.*, 1996; Ashcraft et Liu, 1998). The current fill-in is of the order of $\mathcal{O}(N^2 Log_2 N)$ for 2D finite difference problems and $\mathcal{O}(N^4)$ for 3D problems (Virieux *et al.*, 2009).

1.2.3 Accuracy

The dispersion analysis of the 3D mixed-grid stencil has been developed by Operto *et al.* (2007). Consider an infinite homogeneous velocity model of velocity c and a constant density equal to one. From the appendix C of Operto *et al.* (2007), the discrete wave equation (without PML conditions) reduces to

$$\begin{aligned} & \frac{\omega^2}{c^2}(w_{m1}p_{000} + \frac{w_{m2}}{6}p_1 + \frac{w_{m3}}{12}p_2 + \frac{w_{m4}}{8}p_3) + \frac{w_1}{h^2}(p_1 - 6p_{000}) \\ + \quad & \frac{w_2}{3}\left[\frac{1}{h^2}(p_1 - 6p_{000}) + \frac{1}{4h^2}(2p_2 - 24p_{000})\right] + \frac{w_3}{4}(6p_3 - 4p_2 + 8p_1 - 48p_{000}), \quad (1.27) \end{aligned}$$

where

$$\begin{aligned} p_1 &= p_{100} + p_{010} + p_{001} + p_{-100} + p_{0-10} + p_{00-1}, \\ p_2 &= p_{110} + p_{011} + p_{101} + p_{-110} + p_{0-11} + p_{-101} + p_{1-10} \\ &+ p_{01-1} + p_{10-1} + p_{-1-10} + p_{0-1-1} + p_{-10-1}, \end{aligned}$$

$$p_3 \;=\; p_{111} + p_{-1-1-1} + p_{-111} + p_{1-11} + p_{11-1} + p_{-1-11} + p_{1-1-1} + p_{-11-1}.$$

Following a classic harmonic approach, I insert the discrete expression of a plane wave, $p_{lmn} = e^{-\iota hk(l\cos\phi cos\theta + m\cos\phi\sin\theta + n\sin\phi)}$ where $\iota^2 = -1$, in equation (1.27). The phase velocity is given by ω/k. The normalized phase velocity is defined by $V_{ph} = v_{ph}/c$ and the number of nodes per wavelength λ by $G = \lambda/h = 2\pi/kh$. After some straightforward manipulations, although cumbersome, the following expression for the numerical phase velocity is obtained as

$$V_{ph} = \frac{G}{\sqrt{2J}\pi}\sqrt{w_1(3-C) + \frac{w_2}{3}(6-C-B) + \frac{2w_3}{4}(3-3A+B-C)}, \tag{1.28}$$

where $J = (w_{m1} + 2w_{m2}C + 4w_{m3}B + 8w_{m4}A)$ with

$$\begin{aligned} A &= \cos a\cos b\cos c, \\ B &= \cos a\cos b + \cos a\cos c + \cos b\cos c, \\ C &= \cos a + \cos b + \cos c. \end{aligned}$$

with expressions $a = 2\pi/G\cos\phi\cos\theta$, $b = 2\pi/G\cos\phi\sin\theta$ and $c = 2\pi/G\sin\phi$. We can check that, $V_{ph} \longrightarrow 1$ when $G \longrightarrow \infty$ for $J = 1$ and for the 3 cases $(w_1, w_2, w_3) = (1,0,0), (0,1,0)$ and $(0,0,1)$ whatever ϕ and θ are. This validates the expression of the phase velocity in equation (1.28).

Operto *et al.* (2007) estimated the five independent parameters w_{m1}, w_{m2}, w_{m3}, w_1, w_2 which minimize the least-squares norm of the misfit of the normalized phase velocity $1 - V_{ph}$. They found that the values $w_{m1} = 0.4964958$, $w_{m2} = 7.516875E-02$, $w_{m3} = 4.373936E-03$, $w_1 = 1.8395265E-05$ and $w_2 = 0.890077$, which imply $w_{m4} = 5.69038E-07$ and $w_3 = 0.1099046$. The coefficients show that stencils of types 2 and 3 have a dominant contribution in the mixed-grid stencil. On the other hand, the mass coefficients show a dominant contribution of the coefficients located at the collocation node and at the nodes associated with the stencil of type 1.

The dispersion curves for the three kinds of stencils 1, 2 and 3 without mass averaging are shown in figure 1.3. These stencils used individually would require up to 40 grid nodes per wavelength. Stencils present different behaviours related to their isostropic preferential directions. The phase velocity dispersion curve for the mixed stencil with mass averaging are shown in figure 1.4 and displays a dramatic improvement. We may recommend that 4 grid nodes for wavelength are necessary for neglecting the numerical dispersion as we shall see now.

Brossier *et al.* (2010b) studied the sensitivity of the accuracy of the mixed-grid stencil to the choice of the weighting coefficients w_{m1}, w_{m2}, w_{m3}, w_1, w_2, w_3. They designed an accurate stencil for a discretization criterion of 4 grid-points per minimum propagated wavelength. This criterion is driven by the spatial resolution of full waveform inversion, which is half-a-wavelength. They gave a table 1.1 of the weighting coefficients as a function of G_m. For high values of G_m, the Cartesian stencil has a dominant contribution (highlighted by the value of w_1, while the first rotated stencil has the dominant contribution for low values of G_m as shown by the value of w_2. The fact that the Cartesian stencil is dominant for large values of G_m is consistent with the fact that this stencil has a smaller spatial support (i.e., $2 \times h$) than rotated stencils and that it provides accurate solutions for G greater or equal to 10 (Virieux, 1984). The error on the phase velocity is plotted in polar coordinates for four values of $G_m = 4, 6, 8, 10$

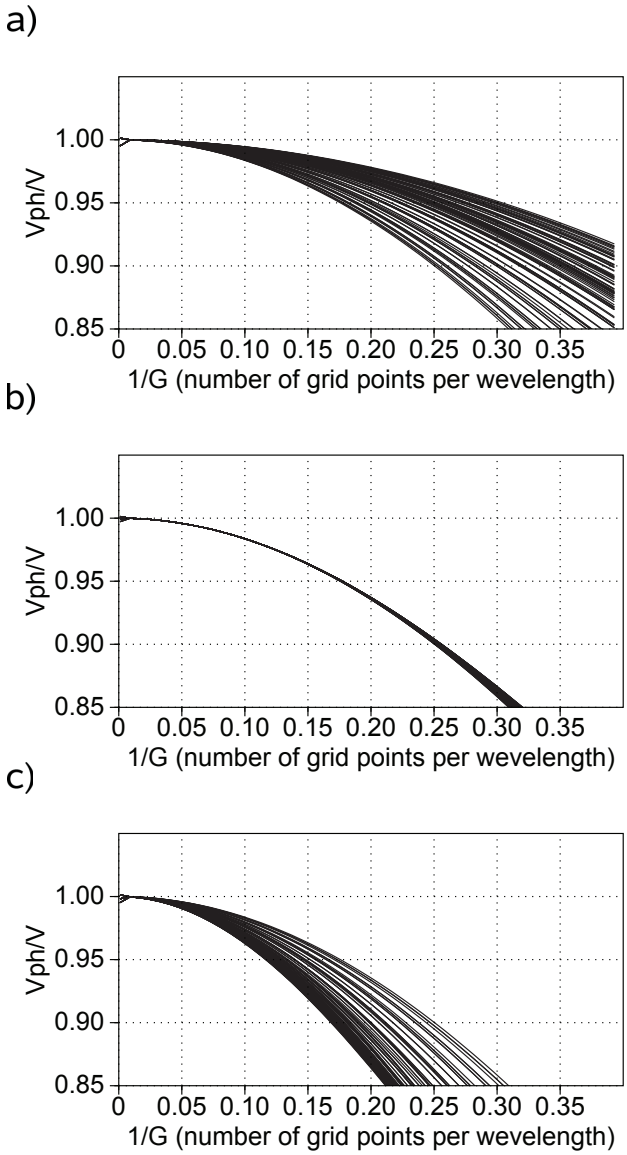

Figure 1.3: Dispersion curves for phase velocity (Operto *et al.*, 2007). (a) Stencil 1 without mass averaging. (b) Stencil 2 without mass averaging. (c) Stencil 3 without mass averaging. The curves are plotted for angles θ and ϕ ranging from 0 to 45^o.

(Figure 1.5a). The phase velocity dispersion is negligible for $G = 4$ (Brossier *et al.*, 2010b). However, more significant error (0.4%) is obtained for intermediate values of G (*i.e.* $G_m = 6$ in the figure 1.5a). This highlights the fact that the weighting coefficients are optimally designed to minimize the dispersion for one grid interval in an homogenous media for all directions. The phase-velocity error is more uniform over directions over these values of G. The maximum of this error is reduced (0.25% *against* 0.4%). However the improved isotropic property of the mixed-grid stencil is degraded and the phase-velocity dispersion is significantly increased for

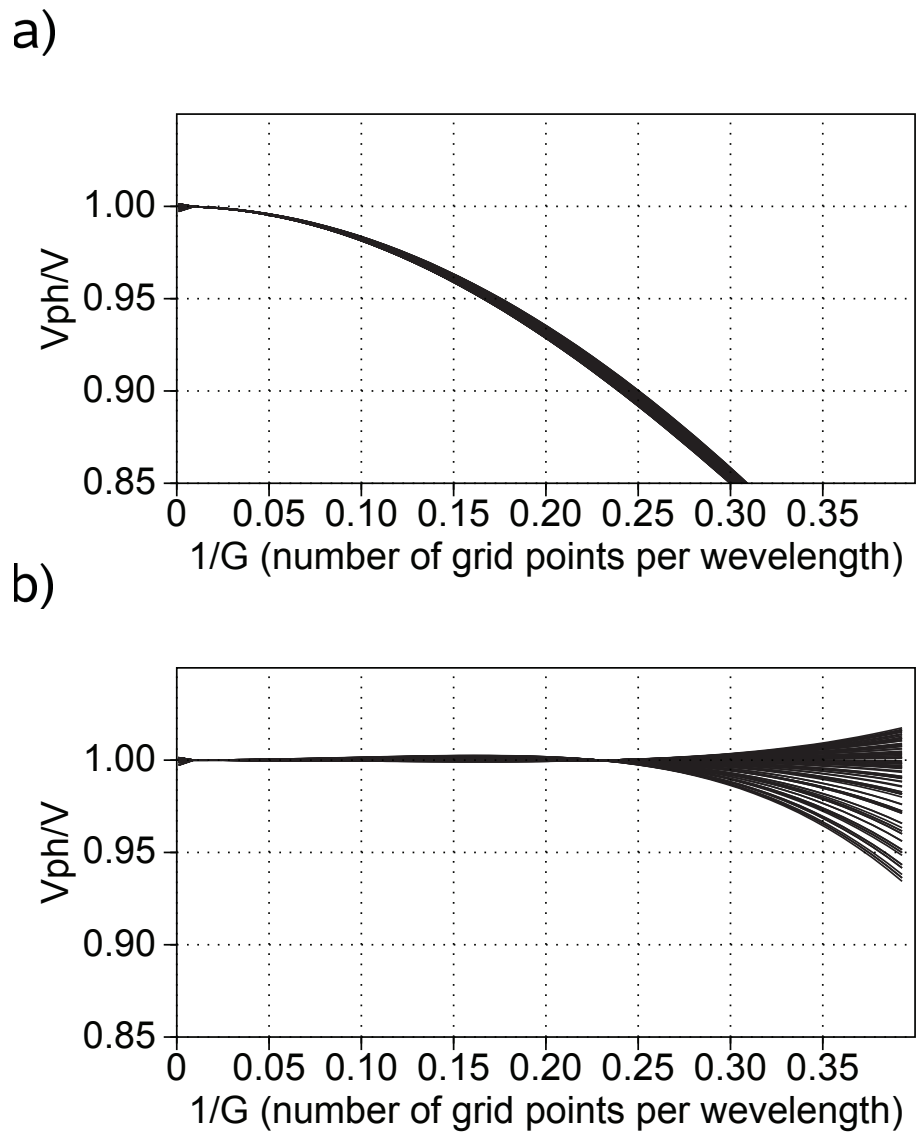

Figure 1.4: Phase velocity dispersion curve for mixed-grid stencil (Operto *et al.*, 2007): (a) without mass averaging. (b) with mass averaging. The curves are plotted for angles θ and ϕ ranging from 0 to 45^o.

$G_m = 4$ (Brossier *et al.*, 2010b).

Table 1.1: Coefficients of the mixed-grid stencil as a function of the discretization criterion G_m for the minimization of the phase velocity dispersion (Brossier *et al.*, 2010b).

G_m	4,6,8,10	4	8	10	20	40
wm_1	0.4966390	0.5915900	0.5750648	0.7489436	0.7948160	0.6244839
wm_2	7.512332E-02	4.965349E-02	5.767590E-02	1.390442E-02	3.713921E-03	5.066460E-02
wm_3	4.384638E-03	5.108510E-03	5.569136E-03	6.389212E-03	5.540431E-03	1.423687E-03
wm_4	6.761402E-07	6.148369E-03	1.506268E-03	1.136992E-02	1.455191E-02	6.80553E-03
w_1	5.024800E-05	8.80754E-02	0.133953	0.163825	0.546804	0.479173
w_2	0.8900359	0.8266806	0.7772883	0.7665769	0.1784437	0.2779923
w_3	0.1099138	8.524394E-02	8.875889E-02	6.959790E-02	0.2747527	0.2428351

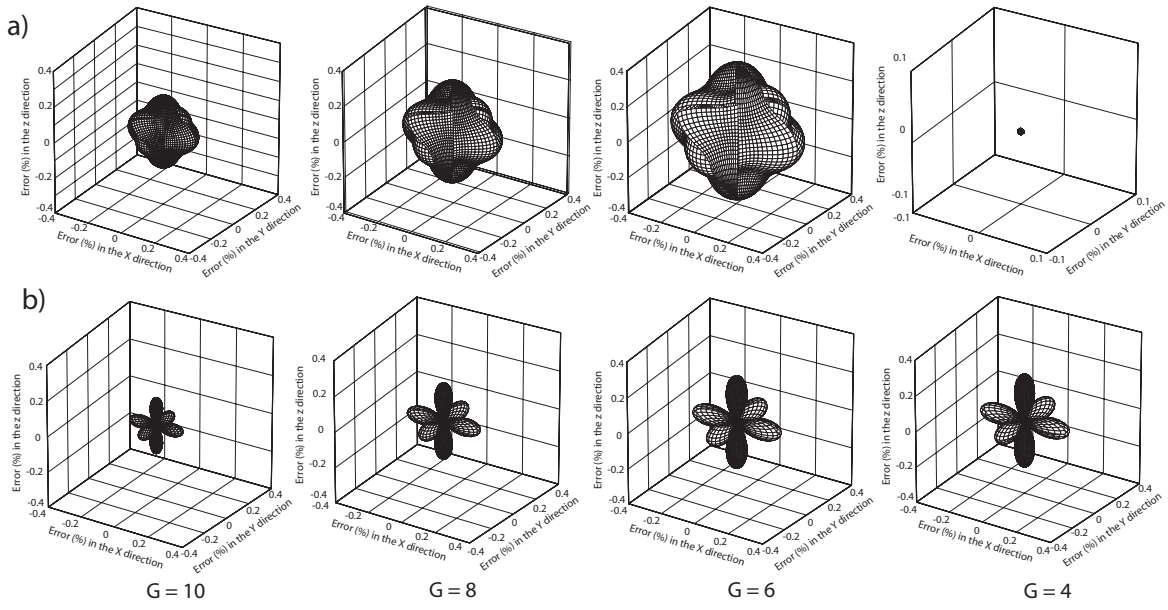

Figure 1.5: Phase-velocity dispersion curves as a function of 1/G. The different curves are associated with the two incidence angles of the plane wave in the 3D homogeneous medium (Brossier *et al.*, 2010b) (a) the phase-velocity dispersion was minimized for 4 values of G: 4, 6, 8 and 10. (b) The phase-velocity dispersion was minimized for G = 4.

1.3 Finite-difference time-domain discretization

1.3.1 Staggered-grid stencil

For crustal scale, the most widely used technique for numerical modeling of seismic wave propagation is the finite-difference method (Virieux, 1986; Levander, 1988; Graves, 1996). Since the memory requirements and computational costs of such simulations are significant although less than for the frequency approach, it is desirable that the coarsest possible finite-difference grid should be used.

The finite difference method belongs to the so-called grid-point methods. In the grid-point methods, a computational domain is covered by a space-time grid and each function is represented by its values at grid points. The most natural choice of the **FD** grid for the displacement formulation is the conventional grid in which all displacement and body-force components are located at each grid point. The conventional grid had been used for many early seismology applications (Alterman et Karal, 1968; Boore, 1972; Alford *et al.*, 1974a; Kelly *et al.*, 1976). Because the conventional-grid displacement **FD** schemes had problems with instabilities in models with high-velocity contrasts and with grid dispersion in media with high Poisson's ratio, Virieux (1984, 1986) introduced the staggered-grid velocity-stress **FD** schemes for modeling seismic wave propagation. Levander (1988) developed this method to the fourth-order for increasing the computational efficiency. Then the staggered-grid **FD** schemes became the dominant type of schemes in the **FD** time-domain (**FDTD**) modeling of seismic wave propagation.

The staggered-grid finite-difference (**SGFD**) method has been developed by Yee (1966) for Maxwell's equation and has been used by Virieux et Madariaga (1982) for modeling the seismic rupture problem. Then, it has been extensively used by Virieux (1984, 1986) to solve the 2D elastodynamic wave equation, especially when the Poisson ration goes towards the value 0.5 which is for fluids.

We will briefly review the methodology of second-order and fourth-order central (**FD**) approximation and staggered-grids. For more details, see Virieux (1984, 1986). And then we shall show the validation of our implementation of the Levander fourth-order scheme in homogeneous media with the analytical solutions.

We now consider three points $i-1$, i and $i+1$ in space on the x axis with the step h, and the discrete fields are u_{i-1}, u_i and u_{i+1} respectively. In accordance with the Taylor's series expansion, if we drop the terms which have the power greater than 4, we have the following equations

$$\begin{aligned} u_{i+1} &= u_i + h\frac{\partial u_i}{\partial x} + \frac{h^2}{2}\frac{\partial^2 u_{i+\epsilon}}{\partial x^2} + \frac{h^3}{6}\frac{\partial^3 u_{i+\epsilon}}{\partial x^3} \\ u_{i-1} &= u_i - h\frac{\partial u_i}{\partial x} + \frac{h^2}{2}\frac{\partial^2 u_{i-\epsilon}}{\partial x^2} - \frac{h^3}{6}\frac{\partial^3 u_{i+\epsilon}}{\partial x^3}. \end{aligned} \tag{1.29}$$

The difference gives

$$u_{i+1} - u_{i-1} = 2h\frac{\partial u_i}{\partial x} + \mathcal{O}(h^3), \tag{1.30}$$

which leads to

$$\frac{\partial u_i}{\partial x} = \frac{u_{i+1} - u_{i-1}}{2h} + \frac{h^2}{3}\frac{\partial^3 u_{i+\epsilon}}{\partial x^3}. \tag{1.31}$$

This approximation is referred to as a second-order accurate, central difference approximation to the first partial derivative of the field u.

Let us now consider a leapfrog scheme on staggered-grids

$$\begin{aligned} u_{i+1/2} &= u_i + h/2\frac{\partial u_i}{\partial x} + \frac{h^2}{4}\frac{\partial^2 u_i}{\partial x^2} + \frac{h^3}{48}\frac{\partial^3 u_i}{\partial x^3} \\ u_{i-1/2} &= u_i - h/2\frac{\partial u_i}{\partial x} + \frac{h^2}{4}\frac{\partial^2 u_i}{\partial x^2} - \frac{h^3}{48}\frac{\partial^3 u_i}{\partial x^3}, \end{aligned} \tag{1.32}$$

which gives

$$\frac{\partial u_i}{\partial x} = \frac{u_{i+1/2} - u_{i-1/2}}{h} + \mathcal{O}(h^2). \tag{1.33}$$

Figure 1.6: Leapfrog second-order accurate central finite-difference approximation of the derivative

This system allows us to discretize the differential operators of the acoustic wave equation (1.9) into leapfrog second-order accurate central-difference scheme (Figure 1.6) on staggered-grids. In this discrete system, the field is known at half-integer positions, while the derivative is known at integer positions.

If we consider now four points for the accusative evaluation, we have as before

$$\begin{aligned} u_{i+1/2} &= u_i + h/2\frac{\partial u_i}{\partial x} + \frac{h^2}{4}\frac{\partial^2 u_i}{\partial x^2} + \frac{h^3}{48}\frac{\partial^3 u_i}{\partial x^3} + \frac{h^4}{384}\frac{\partial^4 u_i}{\partial x^4} + \mathcal{O}(h^5) \\ u_{i-1/2} &= u_i - h/2\frac{\partial u_i}{\partial x} + \frac{h^2}{4}\frac{\partial^2 u_i}{\partial x^2} - \frac{h^3}{48}\frac{\partial^3 u_i}{\partial x^3} + \frac{h^4}{384}\frac{\partial^4 u_i}{\partial x^4} + \mathcal{O}(h^5). \end{aligned} \tag{1.34}$$

The difference is defined as

$$u_{i+1/2} - u_{i-1/2} = h\frac{\partial u_i}{\partial x} + \frac{h^3}{24}\frac{\partial^3 u_i}{\partial x^3} + 8\mathcal{O}(h^5), \tag{1.35}$$

and, for the two other points, we have the values of this wavefield given by

$$\begin{aligned} u_{i+3/2} &= u_i + 3h/2\frac{\partial u_i}{\partial x} + \frac{9h^2}{8}\frac{\partial^2 u_i}{\partial x^2} + \frac{27h^3}{48}\frac{\partial^3 u_i}{\partial x^3} + \frac{h^4}{384}\frac{\partial^4 u_i}{\partial x^4} + \mathcal{O}(h^5) \\ u_{i-3/2} &= u_i - 3h/2\frac{\partial u_i}{\partial x} + \frac{h^2}{4}\frac{\partial^2 u_i}{\partial x^2} - \frac{27h^3}{48}\frac{\partial^3 u_i}{\partial x^3} + \frac{h^4}{384}\frac{\partial^4 u_i}{\partial x^4} + \mathcal{O}(h^5). \end{aligned} \tag{1.36}$$

The difference is given by

$$u_{i+3/2} - u_{i-3/2} = 3h\frac{\partial u_i}{\partial x} + \frac{9h^3}{8}\frac{\partial^3 u_i}{\partial x^3} + \mathcal{O}(h^5). \tag{1.37}$$

We may look for an approximation of the derivation $\partial u_i/\partial x$ as a linear combination given by

$$\frac{\partial u_i}{\partial x} = b_1(u_{i+1/2} - u_{i-1/2}) + b2(u_{i+3/2} - u_{i-3/2}), \tag{1.38}$$

where weights b_1 and b_2 should be found. The following system,

$$\begin{cases} 1. = hb_1 + 3hb_2 \\ 0. = b_1\frac{h^3}{24} + b_2\frac{9h^3}{8}, \end{cases} \tag{1.39}$$

is solved in order to obtain the coefficients. Solutions are $b1 = 9/8h$ and $b2 = -1/24h$. We can obtain the fourth-order accurate central finite-difference approximation to the first partial derivative of u as

$$\frac{\partial u_i}{\partial x} \approx [\frac{1}{h}(u_{i+1/2} - u_{i-1/2}) - \frac{1}{24}(u_{i+3/2} - u_{i-3/2})]. \tag{1.40}$$

By using discretized operators (1.40), a fourth-order accurate in space and second-order in time $\mathcal{O}(\Delta x^4, \Delta t^2)$ discretization of the velocity-stress formulation of the acoustic wave equation (1.9) in three dimensions on a staggered-grid is written as

$$\begin{aligned}
\frac{P^{n+1}_{i,j,k} - P^n_{i,j,k}}{\Delta t} &= k(i,j,k)\frac{b1(Vx^{n+1/2}_{i+1/2,j,k} - Vx^{n+1/2}_{i-1/2,j,k}) + b2(Vx^{n+1/2}_{i+3/2,j,k} - Vx^{n+1/2}_{i-3/2,j,k})}{\Delta x} \\
&+ k(i,j,k)\frac{b1(Vx^{n+1/2}_{i,j+1/2,k} - Vx^{n+1/2}_{i,j-1/2,k}) + b2(Vx^{n+1/2}_{i,j+3/2,k} - Vx^{n+1/2}_{i,j-3/2,k})}{\Delta y} \\
&+ k(i,j,k)\frac{b1(Vx^{n+1/2}_{i,j,k+1/2} - Vx^{n+1/2}_{i,j,k-1/2}) + b2(Vx^{n+1/2}_{i,j,k+3/2} - Vx^{n+1/2}_{i,j,k-3/2})}{\Delta z} + s_{i,j,k} \\
\frac{Vx^{n+1/2}_{i+1/2,j,k} - Vx^{n+1/2}_{i-1/2,j,k}}{\Delta t} &= \frac{b_{i+1/2,j,k}}{\Delta x}[b1\left(P^n_{i+1,j,k} - P^n_{i,j,k}\right) + b2\left(P^n_{i+2,j,k} - P^n_{i-1,j,k}\right)] \\
\frac{Vy^{n+1/2}_{i,j+1/2,k} - Vy^{n+1/2}_{i,j-1/2,k}}{\Delta t} &= \frac{b_{i,j+1/2,k}}{\Delta x}[b1\left(P^n_{i,j+1,k} - P^n_{i,j,k}\right) + b2\left(P^n_{i,j+2,k} - P^n_{i,j-1,k}\right)] \\
\frac{Vz^{n+1/2}_{i,j,k+1/2} - Vz^{n+1/2}_{i,j,k-1/2}}{\Delta t} &= \frac{b_{i,j,k+1/2}}{\Delta x}[b1\left(P^n_{i,j,k+1} - P^n_{i,j,k}\right) + b2\left(P^n_{i,j,k+2} - P^n_{i,j,k-1}\right)].
\end{aligned} \tag{1.41}$$

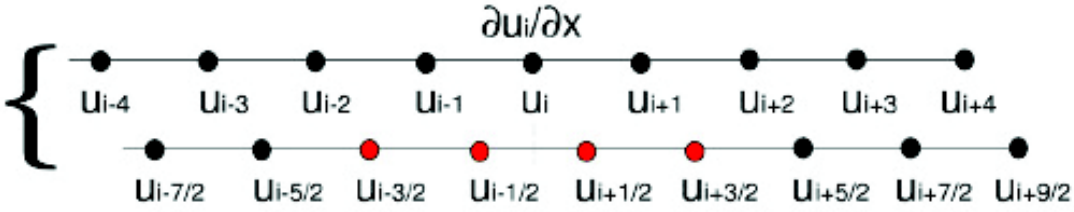

Figure 1.7: Leapfrog fourth-order accurate central-difference approximation

1.3.2 Accuracy and stability

Grid spatial spacing and time step cannot be chosen arbitrarily as a number of conditions should be satisfied to avoid numerical errors and instabilities. A common numerical error in wave propagation modeling occurs when the cell size becomes too large in comparison with the wavelength of the source signal. When this happens, waves undergo dispersion with increasing travel time. This phenomenon is known as grid dispersion. Grid dispersion is a numerical artifact, which causes the higher frequency waves to travel at a different velocity than the lower frequency waves. It has a detrimental effect on the accuracy, the error is controlled by the grid spacing. The question remains as to how fine the spatial grid and which size of the time step should be in order to avoid grid dispersion.

The grid dispersion analysis of the second and fourth-order staggered-grid scheme are given by Alford *et al.* (1974b) and Moczo *et al.* (2000). The dispersive nature of the waveform can be examined by considering phase and group velocity. They gave the expressions of grid phase and group velocities as a function of grid points per wavelength for the second-order staggered-grid scheme in the following (a detail for 4th-order see appendix (Alford *et al.*, 1974b)) as shown by following expressions,

$$\begin{aligned} v_{ph} &= \frac{\omega}{k} = \frac{h}{\pi \Delta t}\frac{\lambda}{h} arcsin\left(C_0 \frac{\Delta t}{h} sin\frac{\pi h}{\lambda}\right) \\ v_g &= \frac{\partial \omega}{\partial k} = \frac{C_0 cos\frac{\pi h}{\lambda}}{\left[1-\left(c_0 - \frac{\Delta t}{h} sin\frac{\pi h}{\lambda}\right)^2\right]^{1/2}}, \end{aligned} \tag{1.42}$$

where $k = 2\pi/\lambda$, C_0 is phase velocity of a homogeneous medium. It is easy to see the dependence of v_{ph} and v_g on a spatial sampling ration h/λ and the $C_0\Delta t/h$ (stability ratio). They are illustrated in figure 1.8 and 1.9. Figures 1.8 and 1.9 show that a discretization rule of 10 grid points per wavelength and 5 grid points per wavelength is repaired for the 2^{nd} order and 4^{th} order accurate stencils.

If the wave propagation modeling algorithm is used as an engine for full waveform inversion, it should be optimal in the sense that the theoretical resolution of full waveform inversion at normal incidence is half a wavelength. If we note the grid interval as h, the maximum wavelength λ_{max}, the $f_{max} = \lambda_{max}/2h$. Now, we can write the resolution with grid spacing as $k_{max} = 2h$. Therefore, we have $\lambda_{max} = 4h$. It means that the scheme $\mathcal{O}(\Delta x^4, \Delta t^2)$ is very well adapted to the expected **FWI** resolution.

For the time step, the Courant-Friedrichs-Lewy condition **CFL** is an other necessary condition for the stability (Courant *et al.*, 1928). It arises when conditionally stable explicit time-marching schemes are used for the numerical solution. As a consequence, the time step must be less than a certain value in many explicit time-marching computer simulations. Otherwise the simulation will produce wildly incorrect results. The general **CFL** condition for the n–dimensional case is expressed as

$$\Delta t \sum_{i=1}^{n} \frac{v_{h_i}}{\Delta h_i} \leq C, \tag{1.43}$$

where v is the velocity, Δt is the time step, Δh is the grid spacing and the coefficient C is a dimensionless constant which depends only on the particular equation to be solved.

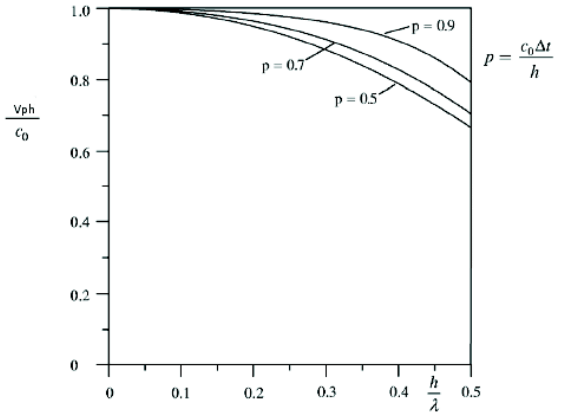

Figure 1.8: Normalized phase velocity for different stability rations for the second-order staggered-grid scheme. After (Alford *et al.*, 1974b).

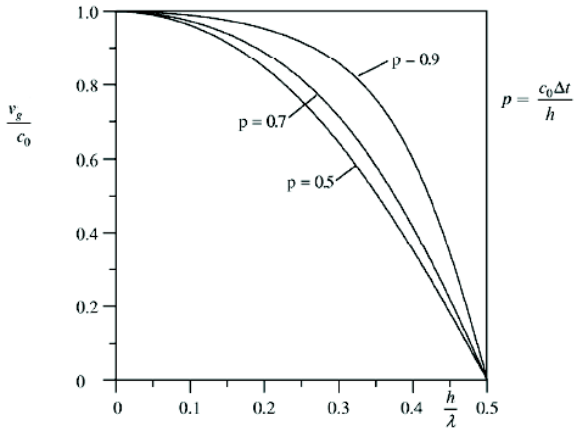

Figure 1.9: Normalized group velocity for different stability rations for the second-order staggered-grid scheme. After (Alford *et al.*, 1974b).

To simplify the problem (1.44), I use

$$\Delta t \leq \xi \frac{\Delta h}{v_{max}}, \tag{1.44}$$

where ξ depends on the scheme. For this work I use $\xi = 0.49$.

Grid dispersion and stability conditions are important features one must consider when doing seismic modeling, the quality of the solution depends on these quantities. To avoid the numerical errors, both the two conditions should be satisfied when we perform a simulation.

We perform a simulation using fourth-order accurately in space and second-order accurately in time. A Ricker signal of central frequency 5 Hz is used for the source. Both grid dispersion and stability conditions are satisfied. The grid interval is adapted to the maximum frequency of the source in order to respect the dispersion condition of four grid points per minimum wavelength. The simulation statistics are summarized in the table 1.2. We propagate around 20 wavelengths. The seismograms are shown in figure 1.10 (a 5 Hz monochromatic wavefield is shown in the chapter 1.3.6). When we fulfill grid dispersion condition and **CFL** condition, we observe a very good agreement with analytical solutions in both time domain (Figure 1.10) and frequency domain (Figure 1.19).

1.3.3 Free surface boundary condition

Realistic simulations require the free surface boundary condition to be implemented. A free surface condition requires to cancel the pressure on the free surface. In this thesis, we are interested in planar free surface because the objective of this thesis is to deal with marine

Table 1.2: The computing statistics: N_x, N_y, N_z are the dimension of the model including PML layers; The time step is denoted by Δt; N_{pml} denotes the number of PML grid; The grid interval Δh is identical in the three dimensions; F denoted the frequency; The physical dimension of the propagating model without PML in the three dimension are denoted by L_x, L_y, L_z.

N_z	N_x	N_y	Δt	Δh	$F(Hz)$	$L_z(km)$	$L_x(km)$	$L_y(km)$	N_{pml}
201	401	401	0.004	50	5	8.0	18	18	20

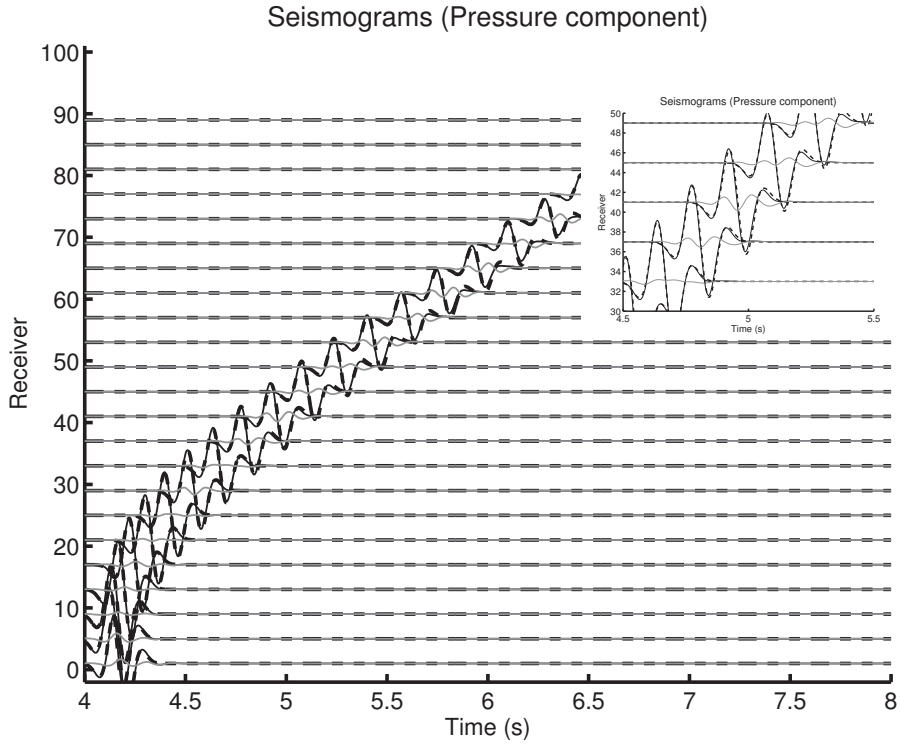

Figure 1.10: Homogeneous infinite space, large offset seismogram. Continuous black line is numerical solution. Dashed black line is the analytical solution. Gray line is the residuals between the numeric and analytical solutions. A zoom part of seismogram from 4.5 s to 5.5 s for receivers $30 \sim 50$ is shown in upper right corner.

application. In the marine case, the surface is relatively planar, therefore, topography should be considered almost flat.

For the acoustic wave equation, there are two ways to implement the free surface boundary condition: first case, the pressure is forced to be zero at the grid points corresponding to the free surface. In this manner, the free surface is aligned along **FD** grid points (on the pressure grid). The second case is the free surface along a virtual plane located half a grid interval above the topside of the **FD** grid (on the velocity grid) (Virieux, 1986). The pressure is cancelled at the virtual free surface by using opposite values of pressure beneath and above free surface.

The two choices of free surface boundary condition can be useful depending on the position

of sources and receivers with respect to node positions. Free surface located at virtual plane is usually used in the cases when the sources or the receivers are very close to the free surface as in marine applications. In this case, the source or the wavefield solution at receiver position will not likely match the node positions and an interpolation function should be used. If we take the simplest case of a linear interpolation in the vertical direction, the solution will be interpolated between the free surface (where the pressure is 0) and the first grid point distant by h for the first case. In the other case, free surface located at virtual grid, the solution will be interpolated between the virtual free surface and the first grid point distant by $h/2$ (Operto, 2006).

We validate our free surface implementation using a $\mathcal{O}(\Delta x^4, \Delta t^2)$ scheme. A Ricker wavelet of central frequency 5 Hz is used as source. The simulation statistics are summarized in table 1.2. The figure 1.11 shows the seismograms, which correspond roughly to 20 propagated wavenumbers (a 5 Hz monochromatic wavefield is shown in chapter 1.3.6 (Figure 1.20)). A very good agreement with analytical solution is shown as for the infinite model case study.

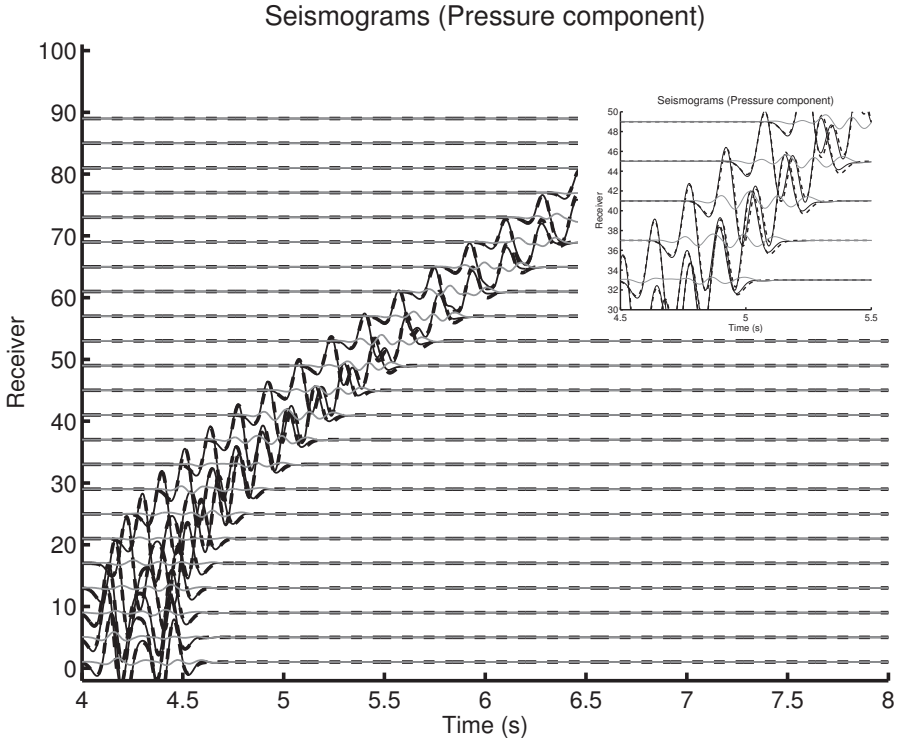

Figure 1.11: Homogeneous half space, large offset seismogram. Continuous black line is numerical solution. Dashed black line is the analytical solution. Gray line is the residuals between the numeric and analytical solutions. A zoom part of seismogram from 4.5 s to 5.5 s for receivers $30 \sim 50$ is shown in upper right corner.

1.3.4 Perfectly-Matched Layers (PML) absorbing boundary conditions

Systems (1.9) and (1.10) define the wave propagation in an infinite medium in time and frequency domains, respectively. However, the wavefield is numerically modeled in a finite domain. Therefore, waves will reflect from the computation edges of the domain, the artificial waves will be recorded. Therefore, we should attenuate or suppress the unwanted waves to simulate a modeling in a infinite medium. Absorbing boundary condition should be designed.

Clayton et Engquist (1977) introduced absorbing boundary conditions based on the paraxial one-way wave equation, while Cerjan *et al.* (1985) introduced the sponge-like absorbing conditions. In this latter case, the computational domain is augmented with sponge layers. The wave propagation is correctly modeled in the computational domain. However, in the sponge layers, the wavefield is progressively attenuated through a damping function. Berenger (1994) improved the Cerjan sponge-like absorbing conditions and proposed the perfectly matched layers (PML) approach. The PML conditions only attenuate the normal component of the wavefield, while the whole wavefield is attenuated in the Cerjan's approach. It has the notable property of having a zero reflection coefficient for all angles of incidence and all frequencies in a continuous medium. However, the perfectly matched layer is not so perfect after discretization, because, in the context of Maxwell's equations and elastodynamic equations, the reflection coefficient is not zero after discretization and even becomes very large at grazing incidence. Kuzuoglu et Mittra (1996) developed the discretized PML at grazing incidence for Maxwell's equation. The main idea is to add to the complex coordinate transform used in classical PML, a frequency-dependent term that implements a Butterworth-type filter in the layer. This modification of the classic PML is called convolutional-PML(C-PML). Komatitsch et Martin (2007) developed an unsplit convolutional perfectly matched layer improved at grazing incidence for the seismic wave equation, which is chosen for this thesis.

For the implementation of the PML conditions in the system (1.9), the pressure wavefield could be split when considering a time approach. The system becomes

$$\begin{aligned}
\frac{\partial p_x(x,y,z,t)}{\partial t} + \gamma_x(x) p_x(x,y,z,t) &= \kappa(x,y,z)\frac{\partial v_x(x,y,z,t)}{\partial x} + s(x,y,z,t) \\
\frac{\partial p_y(x,y,z,t)}{\partial t} + \gamma_y(y) p_y(x,y,z,t) &= \kappa(x,y,z)\frac{\partial v_y(x,y,z,t)}{\partial x} \\
\frac{\partial p_z(x,y,z,t)}{\partial t} + \gamma_z(z) p_z(x,y,z,t) &= \kappa(x,y,z)\frac{\partial v_z(x,y,z,t)}{\partial z} \\
\frac{\partial v_x(x,y,z,t)}{\partial t} + \gamma_x(x) v_x(x,y,z,t) &= b(x,y,z)\frac{\partial p(x,y,z,t)}{\partial x} \\
\frac{\partial v_y(x,y,z,t)}{\partial t} + \gamma_y(y) v_y(x,y,z,t) &= b(x,y,z)\frac{\partial p(x,y,z,t)}{\partial x} \\
\frac{\partial v_z(x,y,z,t)}{\partial t} + \gamma_z(z) v_z(x,y,z,t) &= b(x,y,z)\frac{\partial p(x,y,z,t)}{\partial z},
\end{aligned} \tag{1.45}$$

where the pressure $p(x,y,z,t)$ is split into three unphysical components $p_x(x,y,z,t)$, $p_y(x,y,z,t)$ and $p_z(x,y,z,t)$, and satisfies $p = p_x + p_y + p_z$ (Berenger, 1994).

The 1-D functions γ_x, γ_y and γ_z define the damping functions in the PML layers surrounding the computation medium. These functions differ from zero only inside the PML layers. In the PML layers, we used $\gamma(x) = c_{pml}\ cos(\pi x/2L)$ where L denotes the width of the PML layer and x is a local coordinates in the PML layer, the origin of which is located at the outer edges of

the model. A choice of c_{pml} that minimizes the reflection coefficient in the boundaries, defined in Collino et Tsogka (2001) and optimal for a homogeneous velocity model, is used even for heterogeneous models. The scalar c_{pml} is defined by trial and error depending on the width of the PML layer L. The best value of c_{pml} is the one for which the reflections coming from the edges of the model have the smallest amplitude.

The system of equations is written into the Fourier domain and the functions $\xi_x(x) = 1 + \iota\gamma_x(x)/\omega$, $\xi_y(y) = 1 + \iota\gamma_y(y)/\omega$ and $\xi_z(z) = 1 + \iota\gamma_z(z)/\omega$ are introduced to simplify the equations leading to the system

$$\begin{aligned}
\frac{-\iota\omega\xi_x(x)}{\kappa(x,y,z)}p_x(x,y,z,\omega) &= \frac{\partial v_x(x,y,z,\omega)}{\partial x} + s(x,y,z,\omega)\\
\frac{-\iota\omega\xi_y(y)}{\kappa(x,y,z)}p_y(x,y,z,\omega) &= \frac{\partial v_y(x,y,z,\omega)}{\partial y}\\
\frac{-\iota\omega\xi_z(z)}{\kappa(x,y,z)}p_z(x,y,z,\omega) &= \frac{\partial v_z(x,y,z,\omega)}{\partial z}\\
-\iota\omega v_x(x,y,z,\omega) &= \frac{b(x,y,z)}{\xi_x(x)}\frac{\partial p(x,y,z,\omega)}{\partial x}\\
-\iota\omega v_y(x,y,z,\omega) &= \frac{b(x,y,z)}{\xi_y(y)}\frac{\partial p(x,y,z,\omega)}{\partial y}\\
-\iota\omega v_z(x,y,z,\omega) &= \frac{b(x,y,z)}{\xi_z(z)}\frac{\partial p(x,y,z,\omega)}{\partial z}.
\end{aligned} \tag{1.46}$$

Operto *et al.* (2007) developed the 3D frequency-domain acoustic wave equation with unsplit PML conditions as below

$$\begin{aligned}
\frac{-\iota\omega}{\kappa(x,y,z)}p(x,y,z,\omega) &= \frac{1}{\xi_x(x)}\frac{\partial v_x(x,y,z,\omega)}{\partial x} + \frac{1}{\xi_y(y)}\frac{\partial v_y(x,y,z,\omega)}{\partial y}\\
&+ \frac{1}{\xi_z(z)}\frac{\partial v_z(x,y,z,\omega)}{\partial z} + s(x,y,z,\omega)\\
v_x(x,y,z,\omega) &= \frac{\iota b(x,y,z)}{\omega\xi_x(x)}\frac{\partial p(x,y,z,\omega)}{\partial x}\\
v_y(x,y,z,\omega) &= \frac{\iota b(x,y,z)}{\omega\xi_y(y)}\frac{\partial p(x,y,z,\omega)}{\partial y}\\
v_z(x,y,z,\omega) &= \frac{\iota b(x,y,z)}{\omega\xi_z(z)}\frac{\partial p(x,y,z,\omega)}{\partial z}.
\end{aligned} \tag{1.47}$$

The conventional implementation of perfectly matched layer absorbing boundary condition splits the pressure p into three unphysical acoustic fields p_x, p_y, p_z. These unphysical acoustic fields components are use to account for the PML absorbing boundary conditions (Zhang et Ballmann, 1997; Operto *et al.*, 2002). This split PML method has proved to be very efficient from a numerical point of view for the elastic wave equation to absorb both body waves with non-grazing incidence and surface waves. However, at grazing incidence the split PML method suffers from large spurious reflections that make it less efficient for instance in the case of very thin mesh slices, in the case of sources located at very close to the edge of the mesh, and in the case of receivers located at very large offset (Komatitsch et Martin, 2007). The unsplit PML method can be useful, for instance, in the case of thin mesh slices, source close to edge and

receivers located at very large offset. Komatitsch et Martin (2007) demonstrated the efficiency of unsplit PML method on a thin mesh slice for an isotropic material. In this work, the unsplit PML method is used.

Figure 1.12 shows a study with different PML widths. The snapshot are at the 1.6s propagation, pressure component in the plane XZ. The wavefield is propagated in a homogenous infinite space (V=4000m/s). Size of model is $101 \times 201 \times 201$ ($N_z \times N_x \times N_y$), interval space is 50 m. A Ricker wavelet of central frequency 5 Hz is used as source. We show that the reflection from the PML layers is negligible as long as the width of the PML layer is greater than the dominant wavelength, i.e., the number of grid points in the PML layers are greater than 5 for a $\mathcal{O}(h^4)$ stencil.

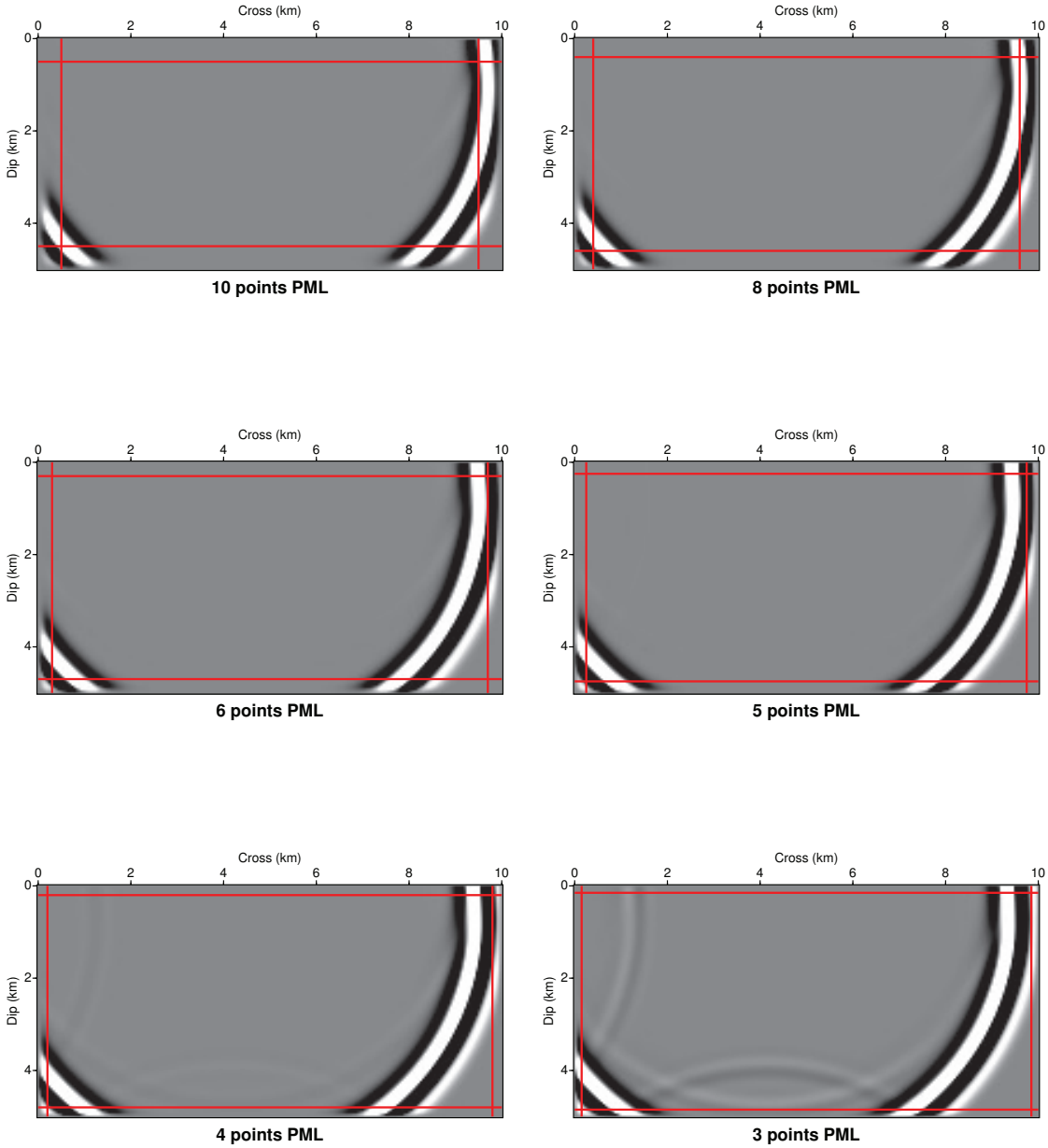

Figure 1.12: Snapshots at 1.6 s of the pressure component in the plane XZ for different PML point definitions . The black point is the source position and the red lines is the limits of the PML. Results are shown with the same amplitude.

1.3.5 Source excitation on coarse grid and extraction of solutions at receiver positions

In finite-difference methods, the source and receiver position are ideally located at the grid nodes. If not, we should initialize the source value to the nearest node called the nearest point approximation. However, the nearest point approximation as well as the linear interpolation are too crude for the required precision in computing traces (Figure 1.13): we get a significant error (RMS = 24.6%). In the case of source or receiver close to the free surface, i.e, the Valhall case, the nearest position for a source located at a depth of 5 m is the free surface where we have exactly a pressure equal to zero. Pressure source is singular at the free surface. Therefore, we consider another interpolation based on a discrete approximation of the Dirac distribution proposed by Hicks (2002). We get an improved result close to the analytic solution. The error reduces to 4.7% (Figure 1.14).

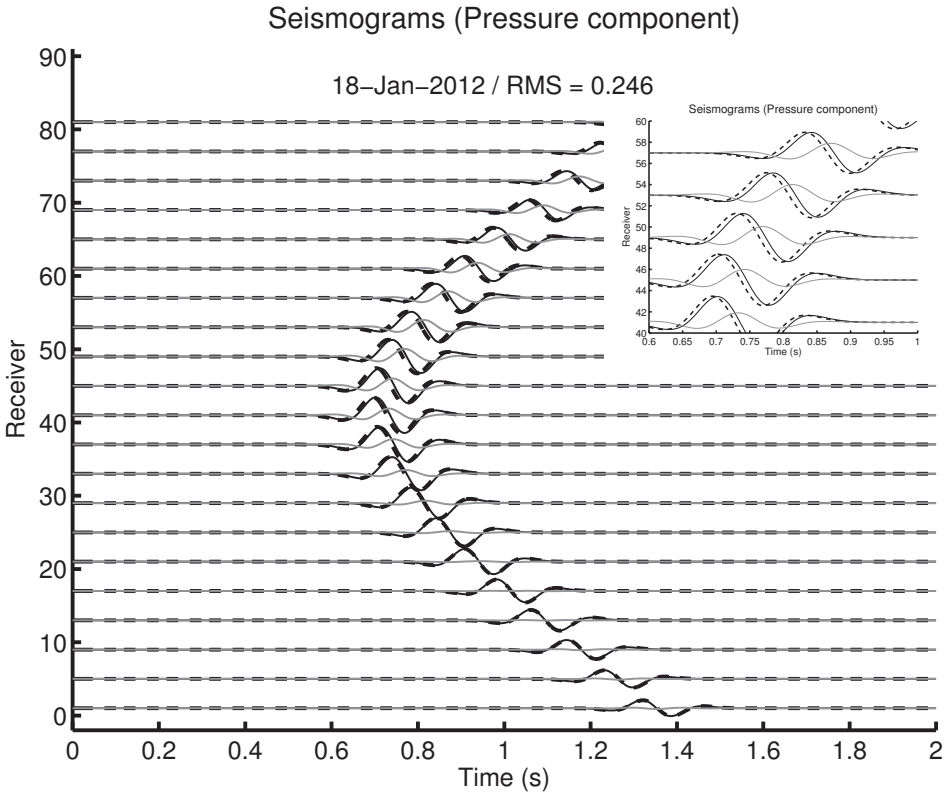

Figure 1.13: Source is implemented at the nearest node, while the source is not located at the grid nodes. In a homogenous (velocity is 4000 m/s) infinite medium, the size of model is $101 \times 201 \times\ 201(n_z \times n_y \times n_x)$, the interval grid is 50 m. The source term is a Ricker wavelet of central frequency 5 Hz located at (525 m 4025 m 4025 m), 81 receivers are located along x at 3500 m depth. We compare here the time solution (pressure component). Continuous black line is numerical solution. Dashed black line is the analytical solution. Gray line is the residuals between the numeric and analytical solutions. We see that the numerical solution is far from the analytical solution both in amplitude and time. A zoom part of seismogram from 0.6 s to 1.0 s for receivers $40 \sim 60$ is shown in the upper right corner.

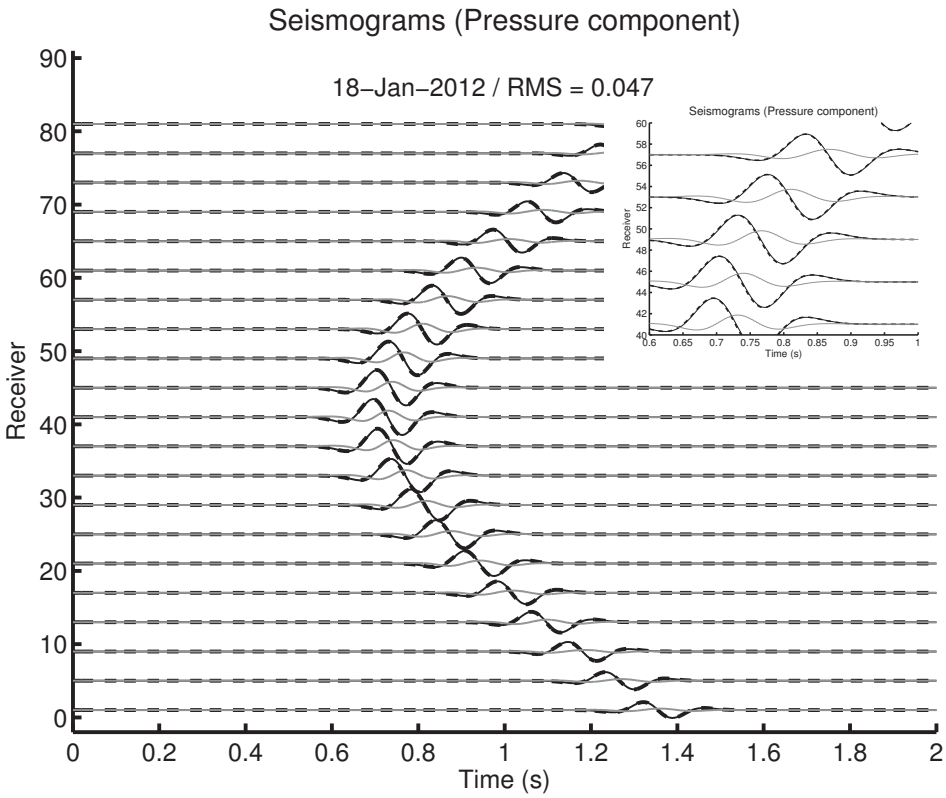

Figure 1.14: *Hicks* interpolation is used. In a homogeneous (velocity is 4000 m/s) infinite medium, the size of model is $101 \times 201 \times 201(n_z \times n_y \times n_x$, the interval grid is 50 m. The source term is a Ricker wavelet of central frequency 5 Hz located at (525 m 4025 m 4025 m), 81 receivers are located along x at 3500 m in depth at the nodes of the grid. We compare here the time solution (pressure component). Continuous black line is numerical solution. Dashed black line is the analytical solution. Gray line is the residuals ($\times$ 5) between the numeric and analytical solutions. We show the improvement of the match between the numerical and analytical solutions compared to that of Figure 1.13. A zoom part is shown in the upper right corner.

The principle of the *Hicks* interpolation is to approximate the spatial distribution of a point source (i.e.,a spatial dirac) with a sinc function, which is windowed by a Kaiser function to limit its spatial support (Figure 1.15). For a point source, the source function is defined as $f(x) = S\delta(x)$, where δ is a spatial delta function, S determines the temporal variation of the source. Then, the source function is sampled on the finite difference grid by

$$f_n = S\delta(n + \alpha), \tag{1.48}$$

where $-0.5 < \alpha \leq 0.5$ and n is an integer, such that $x = n + \alpha$ is the distance from the source in finite-difference nodes. For the source function to be adequately sampled by the finite-difference grid, it is necessary to replace $\delta(x)$ with a band-limited version of itself, $\vec{\delta}(x)$, that has a wavenumber spectrum as close as possible to that of $\delta(x)$ for all wavenumbers less than the Nyquist wavenumber, i.e., for all $k < \pi$, and as close as possible to zero for all $k > \pi$.

Therefore, a sinc function defined by the expression

$$sinc = \frac{\sin(\pi x)}{\pi x}, \tag{1.49}$$

is best suited for our needs: f_n will be nonzero only for $n = 0$ (since $sinc(x) = 0$ for x equal to all nonzero integers). If the source is located between nodes, the discrete source function will be nonzero for all n. For the approximation of a delta function, a spatial window d_n must be applied to $sinc(x)$, so the source function is given by

$$f_n = Sd_n = S\left[W(n+\alpha)sinc(n+\alpha)\right], \tag{1.50}$$

where W represents a windowing function. For a source function with the smallest possible spatial extent, the wavenumber domain properties of d_n are, within the wavenumber range of interest, as close as possible to those of the band-limited $\delta'(x)$. Kaiser proposes a family of simple, yet near-optimal, windowing functions for the design of finite impulse response (FIR) filters (Kaiser, 1974). These windows are given by

$$W_{(x)} = \begin{cases} \frac{I_0(b\sqrt{1-(\frac{x}{r})^2}}{I_0(b))}, & -r <= x <= r \\ 0, & \text{otherwise,} \end{cases} \tag{1.51}$$

where r is half the window width and I_0 is the zero-order modified Bessel function of the first kind. For $b = 0$ the Kaiser window reduces to a rectangular window; as $b \to \infty$, a δ function is obtained. Two useful Kaiser windows are $b = 4.14$ and $b = 6.31$, both with half-widths of $r = 4$ node spacings, and the resulting windowed sinc functions are showed in figure 1.15.

We implement free surface boundary conditions on the top of the model. The nodes located at the free surface correspond to the pressure grid of the finite difference scheme. The pressure component is forced to zero on the grid points located on the free surface. The velocity component of the virtual node above the free surface has the opposite value of the node located below the free surface. We have validated our numerical solutions against analytical solutions. Figure 1.16 shows the case where source and receivers are located at depth in a homogeneous half-space with free surface. The studied case is that both sources and receivers are located closed to free surface. Figures 1.17 and 1.18 show the comparison between the analytical and numerical seismograms when source or receiver is closed to free surface respectively. The model used is a homogeneous (velocity is 4000 m/s) medium, the size of model is $101 \times 201 \times 201(n_z \times n_y \times n_x)$, the interval grid is 50 m. A Ricker signal of central frequency 5 Hz is used as the source term.

The *Hicks* interpolation can be used to locate both source and receiver between finite-difference nodes, and it presents a very good precision even when sources or receivers are close to free surface. We have analyzed the solution of an infinite medium case using a grid spacing of 4 grid points per wavelength. The relative precision of the *Hicks* interpolation solution is 10^{-2} at a distance of 12.5 wavelengths when we compare with the analytical solution, while the relative precision of the nearest node approximation solution is 1. For getting a similar precision to the *Hicks* interpolation, we have to reduce the grid spacing by a factor 10, which lead to increase the computer cost by a factor around $10,000$. These precisions are similar when considering the free surface case. We have analyzed also the *Hicks* interpolation scheme in a parallel environment. The efficiency reduced to 0.975 in comparison with the nearest

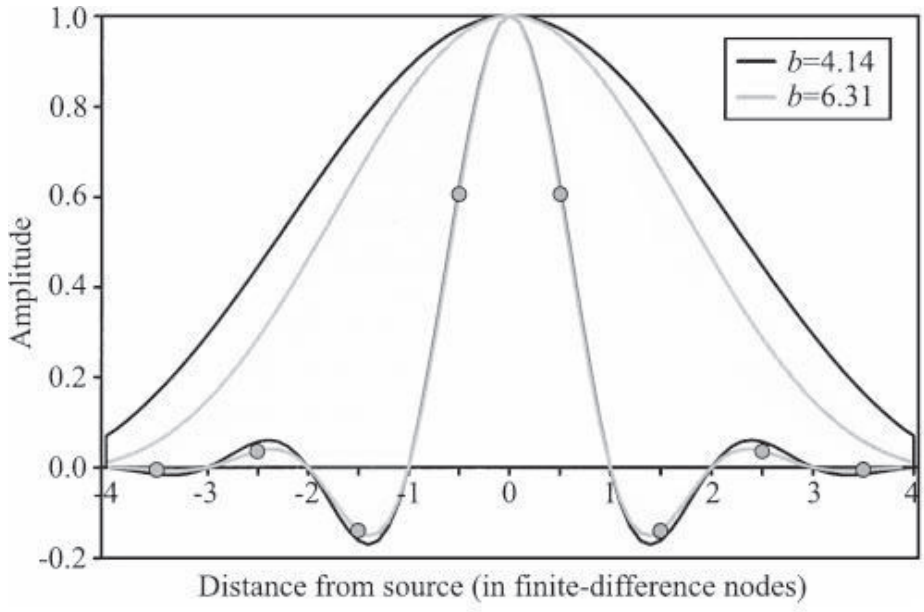

Figure 1.15: Kaiser windowing functions and the corresponding windowed *sinc* functions for two windows with a half-width of $r = 4$ and with windowing parameters $b = 4.14$ (black line) or $b = 6.31$ (gray line). Circles indicate the eight nonzero values of d_n obtained when $r = 4$, $b = 6.31$, and $\alpha = 0.5$. After (Hicks, 2002).

node approximation for interpolating 1 source and 100 receivers, and 0.942 when considering $5,000$ receivers. We foresee that the efficiency will be around 0.91 for the Valhall case ($50,000$ receivers).

1.3.6 Extraction of monochromatic solutions by Discrete Fourier Transform

Due to the time and memory complexities for LU factorization in 3D case frequency domain modeling (Operto *et al.*, 2007), we have chosen the time domain modeling for our frequency FWI algorithm. We should extract the frequency solution to construct the gradient for frequency domain FWI. To extract the frequency domain solution from the time domain modeling, Nihei et Li (2007) and Sirgue *et al.* (2008) proposed an strategy based on Discrete Fourier Transform (DFT) . The advantage of this approach is that the time domain modeling provides the most flexible framework to apply time windowing of arbitrary geometry. Furthermore, DFT allows us to extract an arbitrary number of frequencies within the loop over time steps at minimal extra cost.

The Discrete Fourier Transform is a mathematical procedure used to determine the harmonic, or frequency, content of a discrete signal sequence. Our purpose is to extract monochromatic solutions from modeling in the time domain by DFT. The DFT of a time series signal X_n is written as

$$X_k = \frac{1}{N} \sum_{n=0}^{N-1} x_n \, e^{\frac{-2j\pi kn}{N}} \qquad \forall k = 0, \ldots, N-1, \tag{1.52}$$

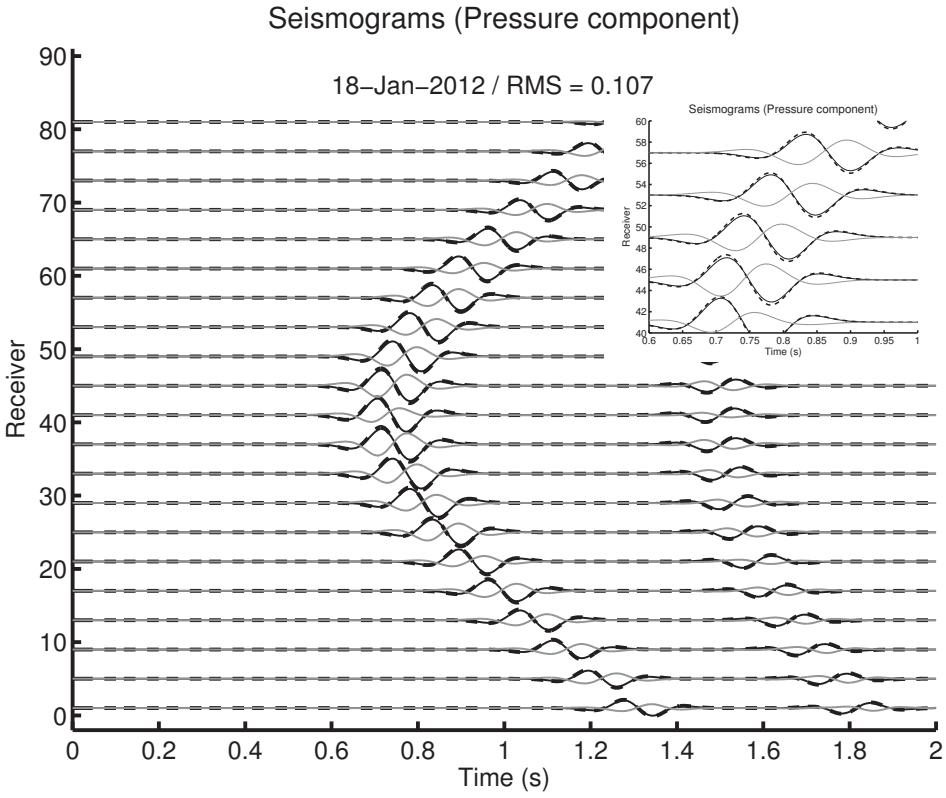

Figure 1.16: Homogeneous half-space, *Hicks* interpolation for source and receivers. Source is at (520 m 4000 m 4030 m) , and 81 receivers are at 3540 m depth according to X with a step 95 m. Time solution (pressure component). Continuous black line is numerical solution. Dashed black line is the analytical solution. Gray line is the residuals ($\times$ 5) between the numeric and analytical solutions. A zoom part of seismogram from 0.6 s to 1.0 s for receivers 40 $\sim$ 60 is shown in the upper right corner.

where n and k are the numbers of samples in the time and frequency series respectively. N indicates the time series (frequency series for the inverse DFT), and $j = \sqrt{-1}$. The inverse DFT is defined as

$$x_n = \sum_{k=0}^{N-1} X_k \, e^{\frac{2j\pi kn}{N}} \qquad \forall n = 0, \ldots, N-1. \tag{1.53}$$

There is an important point here to avoid wraparound. The Nyquist-Shannon Sampling theorem says that a signal can be reconstructed when the sampling rate is more than twice the maximum frequency of the signal being sampled. That means that the highest frequency is given by

$$f_{Nyquist} = \frac{1}{2\Delta t}, \tag{1.54}$$

where Δt is sampling rate in the time domain, therefore the frequency sampling step is written

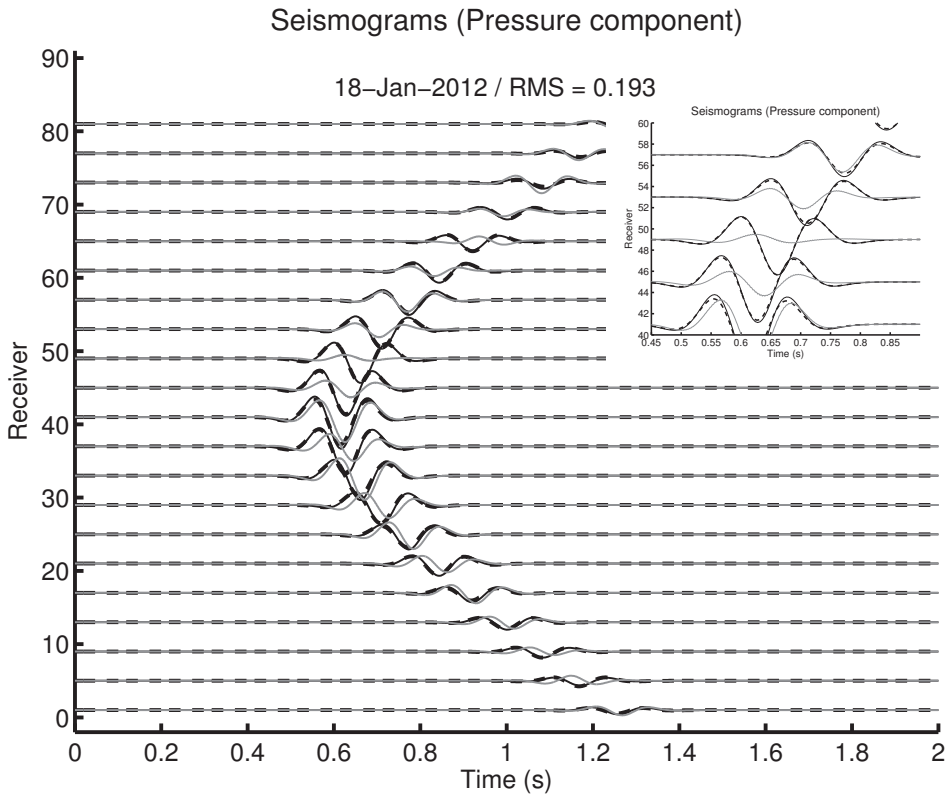

Figure 1.17: Homogeneous half-space, *Hicks* interpolation for source and receivers. Source is at (520 m 4000 m 4030 m), and 81 receivers are at 5 m in depth in X with a step 95 m. Time solution (pressure component). Continuous black line($\times$ 10) is numerical solution. Dashed black line ($\times$ 10) is the analytical solution. Gray line is the residuals ($\times$ 50) between the numeric and analytical solutions. A zoom part of seismogram from 0.6 s to 1.0 s for receivers $40 \sim 60$ is shown in the upper right corner.

as

$$\Delta f = \frac{1}{(N-1)\Delta t}. \tag{1.55}$$

We show here some frequency domain validation tests, which we have been already validated in the time domain. First, we have propagated around 20 wavelengths in an infinite model (Figure 1.19) and a half-space model (Figure 1.20). The simulation statistics are summarized in table 1.2. The frequency solution is extracted at the frequency 5 Hz. Solutions of the two tests show a very good agreement with the frequency domain analytical solution for both the real and imaginary parts. And then, for the different cases with *Hicks* interpolation, we used the simulation statistics are summarized in table 1.3. Source and receiver positions are different with the different tests. The figure 1.21 shows the case of both source and receiver in depth with free surface condition. Figure 1.22 shows the source at the depth and receivers close to the free surface case. The last case the source is located close to free surface and the receivers at the depth. All of these tests show very good agreement with analytical solution. Much more

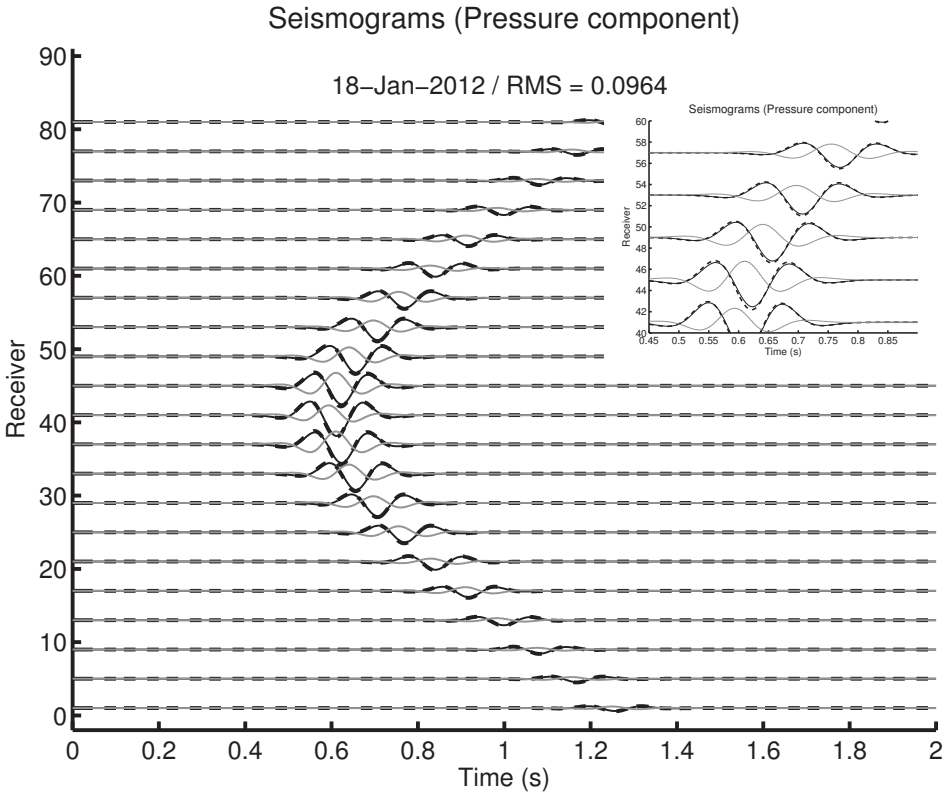

Figure 1.18: Homogeneous half-space, *Hicks* interpolation for source and receivers. Source is at (40 m 4000 m 4030 m), and 81 receivers are at 500 m in depth in X with a step 95 m. Time solution (pressure component). Continuous black line: numerical solution; dashed black line: analytical solution; gray line: residuals ($\times$ 5). A zoom part of seismogram from 0.6 s to 1.0 s for receivers $40 \sim 60$ is shown in the upper right corner.

Table 1.3: The computing statistics: N_x, N_y, N_z are the dimension of the model including PML layers; The time step is denoted by $Deltat$; N_{pml} denotes the number of PML grid; The grid interval Δh is identical in the three dimensions; F denoted the frequency; The physical dimension of the propagating model without PML in the three dimension are denoted by L_x, L_y, L_z.

N_z	N_x	N_y	$Deltat$	Δh	$F(Hz)$	$L_z(km)$	$L_x(km)$	$L_y(km)$	N_{pml}
101	201	201	0.004	50	5	3	8	8	20

precise compared to the nearest point approach validates our *Hicks* interpolation and the DFT implementation.

The major advantage of the DFT is that an arbitrary number or frequencies can be extracted within the loop over the time steps, therefore it is not necessary to store the time series. And also, computing the DFT is not expensive: extraction of multiple frequencies simultaneously does not add significant extra computational cost. Therefore, this approach is particularly

interesting when large number of frequencies N_f must be considered in the inversion in a single iteration. In this case, the number of operations for the explicit method remains unchanged and cost $\mathcal{O}(n^4 N_s)$, but it requires N_f LU factorization for the direct solver. So the number of operations is at the order of $\mathcal{O}(n^6 N_f)$. In the other hand, DFT allows us to extract the frequency solution used for frequency domain FWI from the time domain modeling. Time windowing can be easily applied in the time domain, which is not the case when the modeling is performed in the frequency domain. Time windowing is useful to mitigate the non-linearity of the inversion by extracting specific arrivals (i.e. early arrivals, reflections).

Another interesting point of the explicit scheme is that the seismic acquisition is performed in a defined time window and, therefore, the extraction of frequencies by DFT respects the length of the signal. This is not the case when the system is solved with a direct solver, which provides the steady state solution, equivalent to an infinite time. Therefore, certain phenomena can be found in the solution of the direct solver even though they were not recorded by the seismic acquisition. We note that it is possible to introduce a damping function in the frequency system, equivalent to an exponential decay in time (Shin *et al.*, 2002), but the time window is more delicate to estimate in the approach based on time explicit scheme.

All the validation tests show very good precisions both in amplitude and in phase, for the real and imaginary parts. We have considered the monochromatic wavefield generated by an explosive source in a homogeneous medium (v = 4000 m/s). The duration of simulation is 2 s, and a Ricker signal of central frequency 5 Hz is used as the source function. We obtained an error less than 26% (L_2 norm) compared to analytical solution even in the worst case (sources and receivers close to the free surface), which pose an infinite error without interpolating *hicks* and 19.3% in time domain. By using Discrete Fourier Transform, we can completely reconstruct signals in frequency.

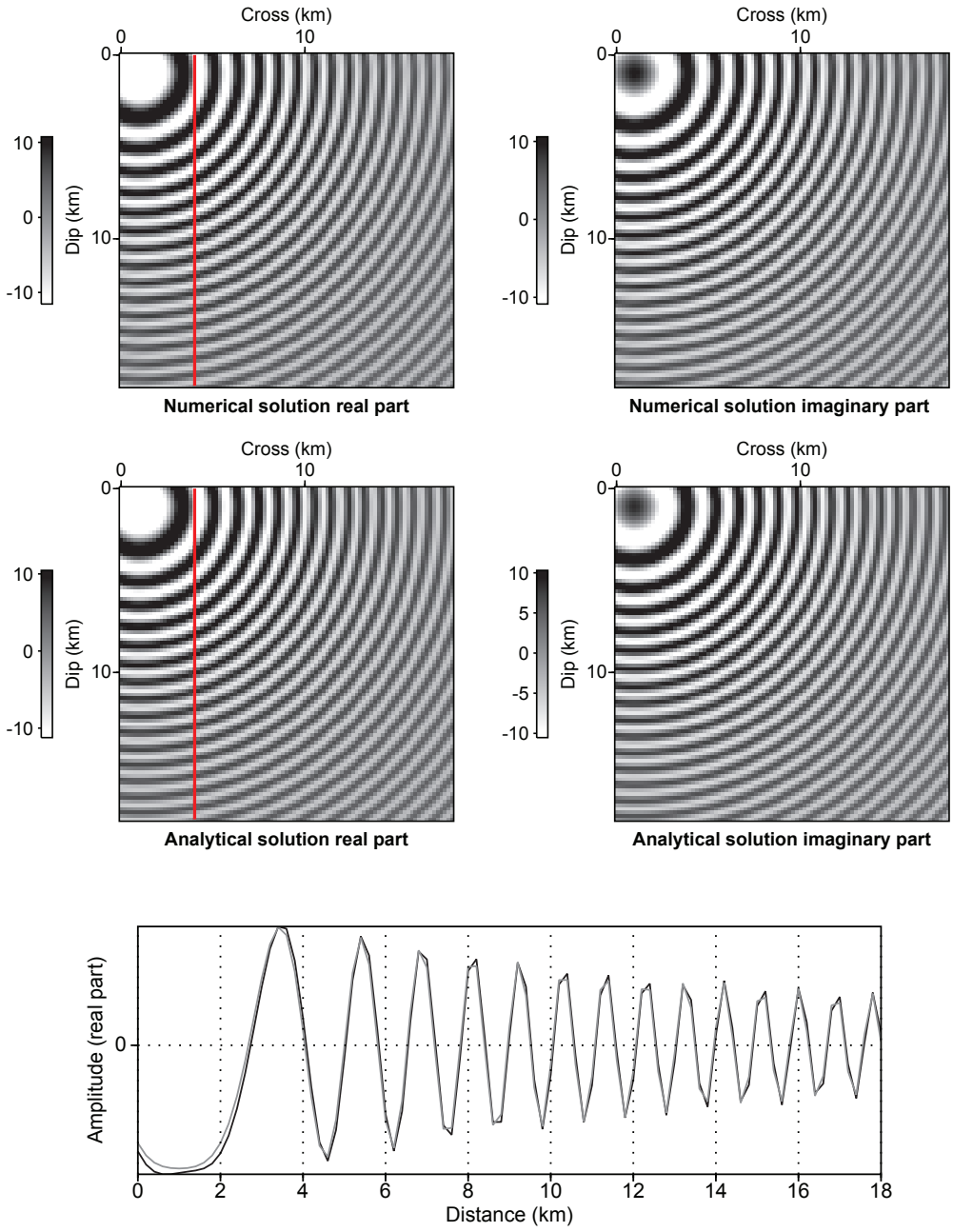

Figure 1.19: Five Hz monochromatic wavefield in a homogeneous infinite space. Top, numerical solutions real and imaginary parts. Middle, analytical solutions real and imaginary parts. Bottom, Real part comparison between numerical(black) and analytical (gray) solutions at receiver positions.

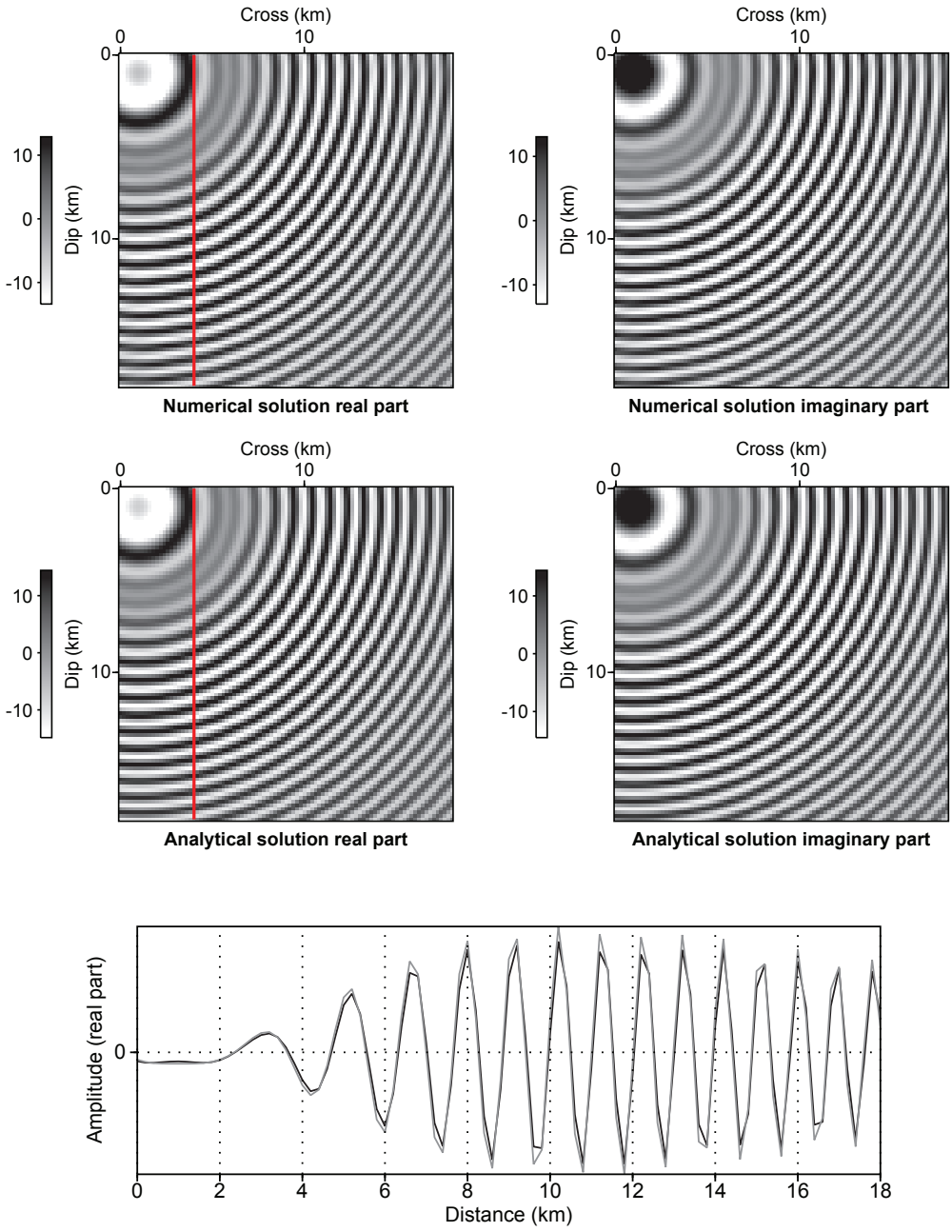

Figure 1.20: Five Hz monochromatic wavefield in a homogeneous half space. Top, numerical solutions real and imaginary parts. Middle, analytical solutions real and imaginary parts. Bottom, Real part comparison between numerical(black) and analytical (gray) solutions at receiver positions.

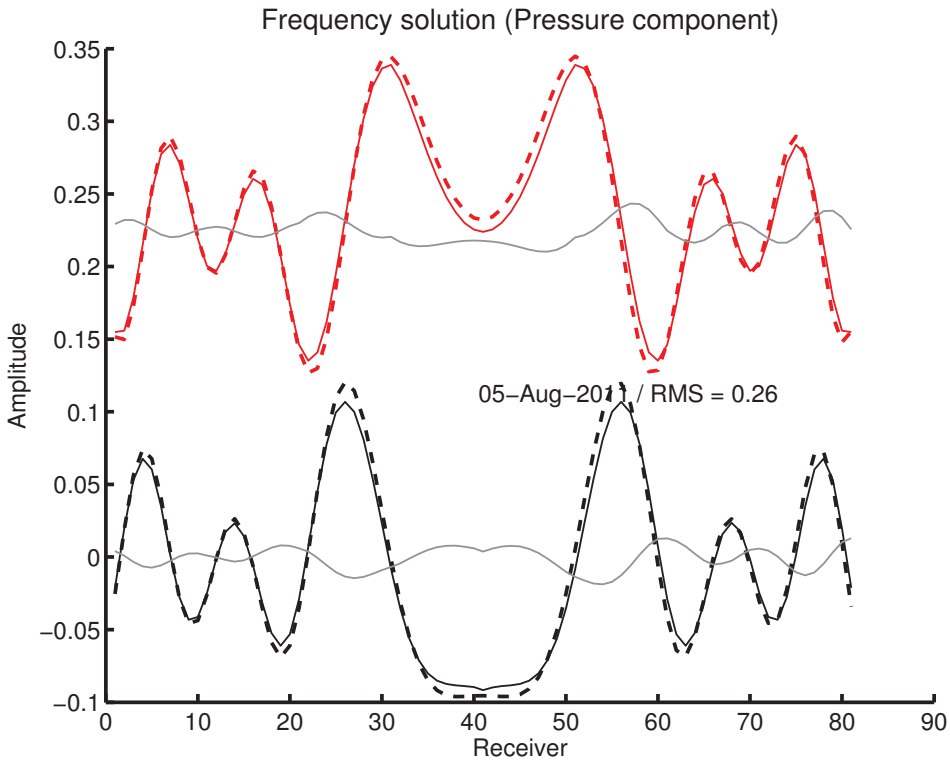

Figure 1.21: Homogeneous half-space, source at (520 m 4000 m 4030 m), and 81 receivers at a depth of 3540 m in X direction. Frequency solution (pressure component). Continuous line is numerical solution; dashed line is the analytical solution; gray line is the residuals between the numerical and analytical solutions. Top, reel part. Bottom, imaginary part.

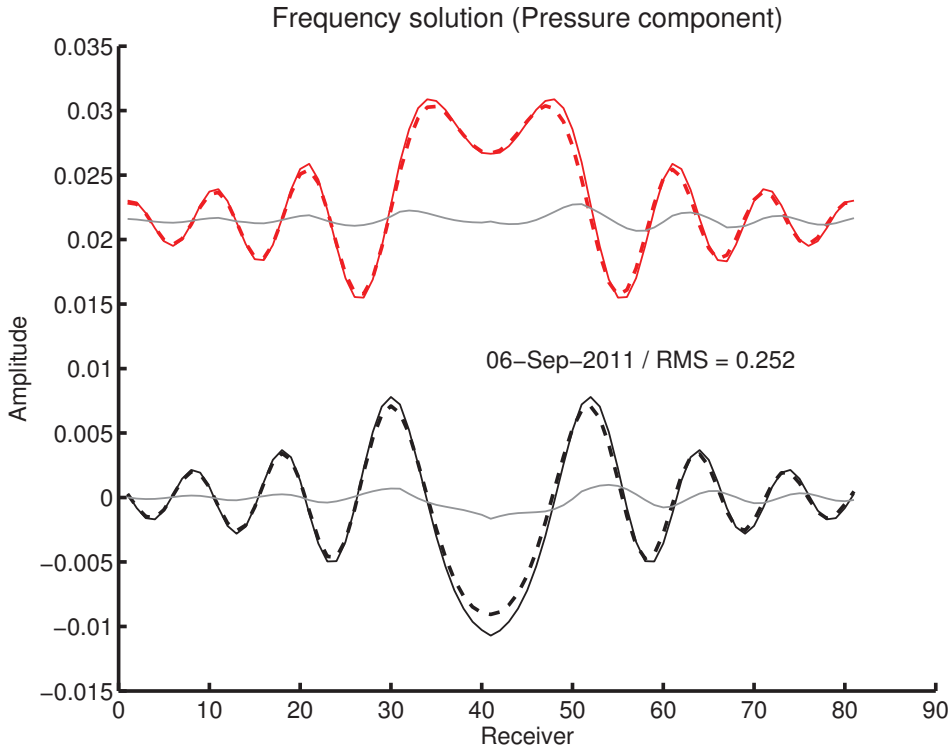

Figure 1.22: Homogeneous half-space, source at (520 m 4000 m 4030 m), and 81 receivers at a depth of 5 m in X direction. Frequency solution (pressure component). Continuous line is numerical solution; dashed line is the analytical solution; gray line is the residuals between the numerical and analytical solutions. Top, reel part. Bottom, imaginary part.

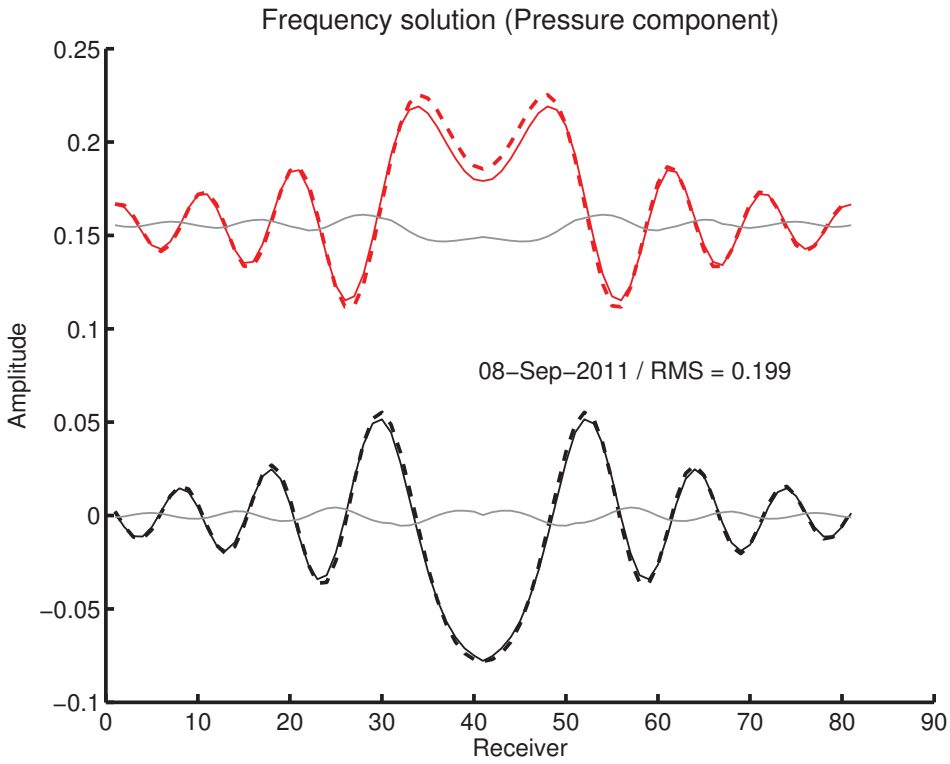

Figure 1.23: Homogeneous half-space, source at (40 m 4000 m 4030 m), and 81 receivers at a depth of 3540 m in X direction. Frequency solution (pressure component). Continuous line is numerical solution; dashed line is the analytical solution; gray line is the residuals between the numerical and analytical solutions. Top, reel part. Bottom, imaginary part.

1.4 Parallel implementation by domain decomposition

1.4.1 Methodology

FWI algorithm must be implemented in parallel to address large scale 3D problem. Depending on the different forward problem approach, different parallel strategies can be considered. The most conventional approach to design parallel time-domain modeling codes for wave propagation rely on domain decomposition methods. The main idea of parallel implementation through domain decomposition of the physical computational domain is to split the original domain of computation into subdomains. Local solutions are computed and needed data are communicated from and by the neighboring domains. The size of subdomain interface has to be much smaller than the size of the overall problem meaning the cost of communicating data is usually almost negligible compared to the computational cost of overall problem. This approach has the advantage of leading itself well to the use of local memory to solve the memory-expensive problem especially in 3D FWI (Brugeas, 1996).

A parallel version of the general algorithm based on the principle of domain decomposition that is suitable for structured meshes is as follows

1. decompose the mesh into subdomain and assign each subdomain to one MPI process;
2. determine its neighboring subdomains for each subdomains;
3. loop of time:
   (a) exchange messages among interfaces;
   (b) calculate the solution.

The physical domain (dimension $n1, n2, n3$) is split in subdomains in each direction specified by the user. The number of grid points ($n1loc, n2loc, n3loc$) in each subdomain is determined by dividing the total number of grid points (PML points included) by the number of subdomains ($nd1, nd2, nd3$) in that direction and assigning the excess final points in the latter subdomains. Each process is identified by its local coordinates, which can either be a single number that corresponds to the $nd1 \times nd2 \times nd3$ coordinate system or three numbers, which correspond to the (z, x, y) coordinate systems. There are simple formulas used to switch back and forth between these two coordinate systems.

The assignment of subdomain to processes is well-ordered, governed by the indexing illustrated below (Figure 1.24). This makes determining the process handing the neighboring subdomains of a current subdomain easy. The current process determines its neighbors on all edges, and the MPI function defines data blocks to be used in sending and receiving messages. The communication is done within the loop in time since the data between interfaces need to be communicated at each time step. The process sends the messages of interface to its neighbors and receives the needed data simultaneously.

The processes should note all of their neighbors. For each process, the six possible neighbors (Up, Down, Left, Right, Front and Behind) are shown in figure 1.25. The three-coordinate indexing make checking for neighbors quite convenient, as subdomains that contain a part of the boundary of the whole domain would either have, in any of the three directions, either one or the maximum number of subdomains in that direction as index value.

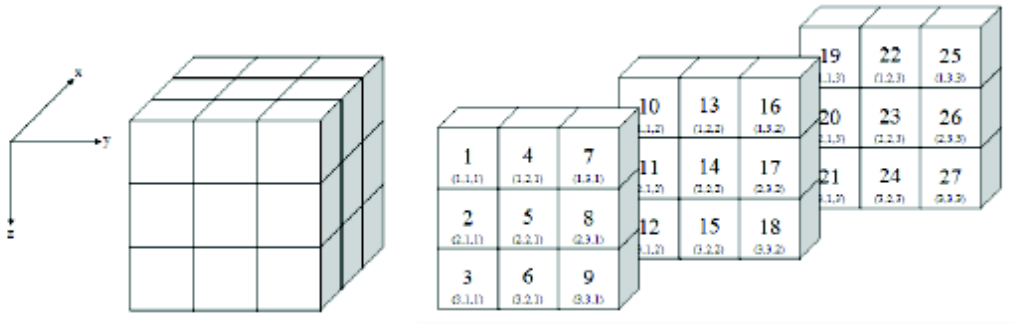

Figure 1.24: Example of indexing processes

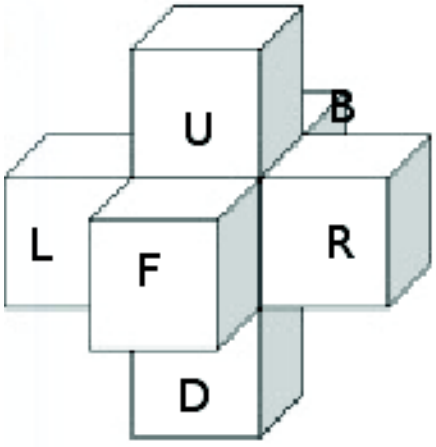

Figure 1.25: The possible neighbors of a subdomain. U, up; D, down; L, left; R, right; F, font; B, behind.

After checking the neighbors, the processes should know the data blocks with which they want to communicate. The MPI procedure typage defines data blocks to be used in communicating, sending, and receiving the messages from its neighbors. MPI offers the advantage to optimize the communications by allowing the user to define customized data types that can be simultaneously sent and received. This saves the user from hard-coding the exact extent of the array needed to be sent and where it would be received. With customized data types, one only needs to specify the initial address of that certain block of data needed. To exchange the data from the overlaps of subdomains, three different types of interface are used in our case (Figure 1.26). To construct the face data types, the MPI function will construct the base type according to z, the fast index firstly, and then x using MPI_TYPE_VECTOR. The stride between these elements are $(n1loc + 4)$*$(size\ of\ data)$, as there are $n1loc$ grid points for the z direction and 2 more additional grids in each side, due to the domain overlap resulting from 4^{th} order accurate stencil. The type face is then constructed from the latest base type with the corresponding number of elements in y direction. The stride between these elements are

Table 1.4: Summarize of the communication

SEND-RECEIVE	SEND ADDRESS	RECEIVE ADDRESS	TYPAGE
Send to UP from DOWN	$x(1,1,1)$	$x(n1loc+1,1,1)$	type_face23
Send to DOWN from UP	$x(n1loc+1,1,1)$	$x(-1,1,1)$	type_face23
Send to RIGHT from LEFT	$x(1,n2loc+1,1)$	$x(1,-1,1)$	type_face13
Send to LEFT from RIGHT	$x(1,-1,1)$	$x(1,n2loc+1,1)$	type_face13
Send to FONT from BEHIND	$x(1,1,-1)$	$x(1,1,n3loc+1)$	type_face12
Send to BEHIND from FONT	$x(1,1,n3loc+1)$	$x(1,1,-1)$	type_face12

$(n1loc+4)$*$(n2loc+4)$*$(size\ of\ data)$.

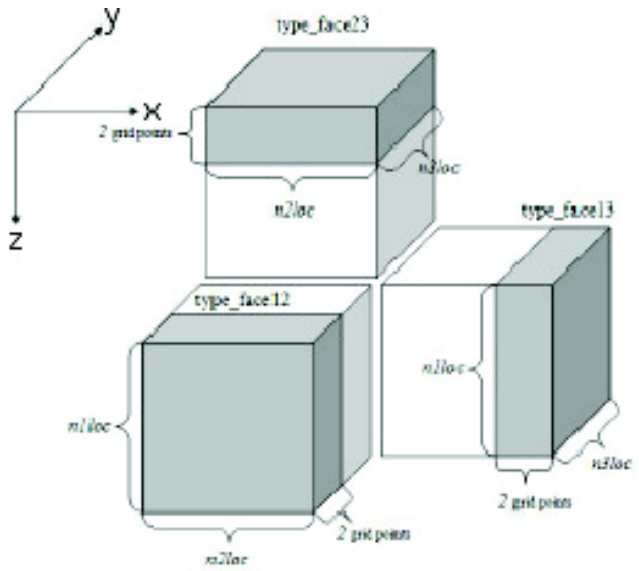

Figure 1.26: The three different types of interface.

The communication is done within the time loop since the data between interfaces need to be communicated at each calculation by using MPI_SENDRECV. The purpose of this procedure is to send data blocks from the subdomain to the corresponding neighboring areas and to receive the same points in the relevant fields. For each subdomain, a three-dimensional array (x) is allocated as $x(-1:n1loc+2,-1:n2loc+2,-1:n3loc+2)$. The original extents of the subdomain $(n1loc,n2loc,n3loc)$ are augmented with 2 more grid points on each side in all the three directions to adapter the need of the 4^{th} order accurate stencil. The procedure of the communication is shown in tableau 1.4.

1.4.2 Scalability

The key issue in parallel processing of a single application is the speedup achieved, especially its dependence on the number of processors used and the proportion of the interface. The

Table 1.5: Decomposition of subdomains in three dimensions

Number of processors	$Z \times Y \times X$
1	$1 \times 1 \times 1$
2	$2 \times 1 \times 1$
4	$1 \times 2 \times 2$
8	$2 \times 2 \times 2$
16	$4 \times 2 \times 2$
32	$2 \times 4 \times 4$

speedup(S) is defined as the factor by which the execution time for the application changes with the number of processors,

$$S = \frac{T_{seq}}{T_N}, \tag{1.56}$$

where T_{seq} is the sequential execution time and T_N is the execution time for N processors.

The figure 1.27 shows a simple test on the FRIPP[1]. The decomposition of subdomains in the three dimensions is shown in the table 1.5.

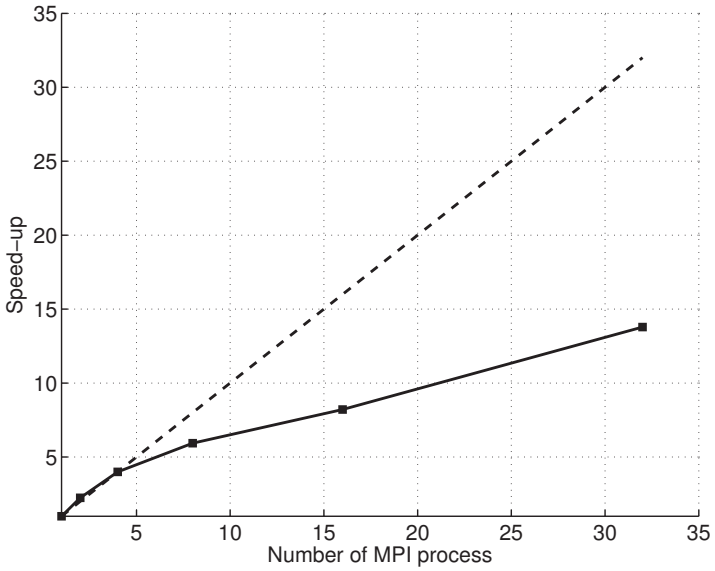

Figure 1.27: The speedup for parallelism of domain decomposition

By assessing the graphs (Figure 1.27), the code does not seem to scale very well. The speedup does not like as theory increasing constant with number of processors (dashed line in figure 1.27). The explanation for this is not yet clear, however it is possible that the way of decomposed domain is one the factors which is summarized on the table given above. Or the proportion of the interface plays a important role (in this test we used a model ($201 \times 201 \times 201$)).

[1]This mid-sized computing center hosted at O.C.A is the result of an joint effort initiated in 2004. 200 cores running at 2.4Ghz Opterons connected via infiniband, and 94 knots IBMx3455 to 8 cores per node AMD Opteron 2.3 GHz.

Table 1.6: Memory and time complexities of DSM HSM, and TDM for multisouce problems and two levels of parallelism: N denotes the dimension of a 3D computational grid; N_p is the total number of processors; N_{rhs} denotes the total number of sources; N_{DSM} denotes the number of MPI processes assigned to LU decomposition in DSM; S_{LU} denotes the speedup of LU factorization; S_s denotes the speedup of substitution step in DSM; N_{HSM} is the number of MPI processes assigned to HSM domain decomposition; k denotes the number of subdomains in one direction and N_{GMRES} denotes the number of GMRES iterations.

Complexity	DSM	HSM	TDM
Memory	$\mathcal{O}(N^4) + \mathcal{O}(N^3 N_{rhs})$	$\mathcal{O}(N^4/k) + \mathcal{O}(N^3 N_{rhs})$	$\mathcal{O}(N^3 N_{rhs})$
Time	$\frac{\mathcal{O}(N^6)}{S_{LU}} + \frac{\mathcal{O}(N^4 N_{rhs} N_{DSM}/N_p)}{S_s}$	$\mathcal{O}(N^4 N_{GMRES} N_{rhs}/k/N_p)$	$\mathcal{O}(N^3 N_t N_{rhs}/N_p)$
Time ($N_{rhs} = N_p$)	$\frac{\mathcal{O}(N^6)}{S_{LU}} + \frac{\mathcal{O}(N^3 N_{DSM})}{S_s}$	$\mathcal{O}(N^4 N_{GMRES}/k)$	$\mathcal{O}(N^3 N_t)$

1.5 Scalability and complexity analysis of time-domain and frequency-domain approaches

We discussed time-domain parallelism using standard domain decomposition methods in the section 1.4, and we have shown the efficiency of these algorithms in the figure 1.27. In the framework of multiple source simulations, the low memory requirements of time-domain methods allow a coarse-grain parallelism over sources (we will discuss details in the section 2.2) that can be combined with domain decomposition parallelism. In the following discussion, the dimension of 3D N^3 computational grid is denoted as N. We consider here only a parallelism over sources. The table 1.6 gives the memory and time complexities for DSM, HSM and TDM approach. Realistic 3D surveys require a significant amount of memory to store the N_{rhs} wavefields distributed over the processors, $\mathcal{O}(N^3 N_{rhs}) = \mathcal{O}(N^5)$, where the number of sources N_{rhs} is over one side of the grid.

Parallelism in DSM is implemented through the use of massive parallel direct solver, such as the software MUMPS (http://mumps.enseeiht.fr/). After parallel lower-upper triangle matrix factorization (LU), the LU factors remain distributed over processors, allowing us to also perform the solution step in parallel. The most limiting factor of direct solvers is their intrinsic limited scalability. Virieux *et al.* (2009) suggested that a speedup greater than 15 is difficult to achieve whatever the number of processors used for both 2D and 3D applications. The memory and time complexities of the DSM for 2D finite-difference problems are $\mathcal{O}(N^2 Log_2 N)$ and $\mathcal{O}(N^3)$, respectively, and increase dramatically in 3D case as $\mathcal{O}(N^4)$ and $\mathcal{O}(N^6)$.

1.6 Modeling in realistic models

We present three-dimensional (3D) simulations performed in the SEG/EAGE Overthrust and Salt models, and we validated the monochromatic wavefield solutions computed with the time-domain approach staggered-grid $\mathcal{O}(\Delta x^4, \Delta t^2)$ against the solutions computed with the direct solver approach. Monochromatic wavefields are extracted from the time-domain solutions using a discrete Fourier transform computed within the time loop (Sirgue *et al.*, 2008).

We implemented the source S in the pressure field for both direct-solver-modeling (**DSM**) and time-domain-modeling (**TDM**) solutions at the same position using the same model for the simulation. The finite-difference time-domain (**FDTD**) code is based on a first-order velocity stress formulation, while the **DSM** finite-difference frequency-domain (**FDFD**) code relies on a second order formulation for pressure. In order to compare the numerical solution of both approaches, it is necessary to implement the source excitation in the two codes consistently.

Let us evaluate the system from time-domain to frequency-domain to find the relation of source term between the two solution. In time-domain we have the following system

$$\begin{aligned} \frac{\partial P}{\partial t} &= k\frac{\partial V}{\partial x} + S \\ \frac{\partial V}{\partial t} &= b\frac{\partial P}{\partial x}, \end{aligned} \tag{1.57}$$

where, P is the pressure, V is wavefield, $k = \rho v^2$ is the bulk modulus, $b = 1/\rho$ is buoyancy and S is the source term. We can write the system (1.57) as the form

$$\begin{aligned} -i\omega P &= k\frac{\partial V}{\partial x} + S \\ -i\omega V &= b\frac{\partial P}{\partial x}. \end{aligned} \tag{1.58}$$

Injecting the second equation of the system (1.58) into the first one, we have

$$(\frac{\omega^2}{k} + \bigtriangledown)P = \frac{i\omega}{k}S. \tag{1.59}$$

It means that, if we implement source term S to the time-domain approach, for the equal solution with frequency-domain solution, we should implement a source term $\frac{i\omega}{k}S$.

1.6.1 The EAGE/SEG Overthrust model

The 3D SEG/EAGE Overthrust model is a constant density acoustic model covering an area of 20 $km \times 20$ $km \times$ 4.65 km , figure 1.28 (Aminzadeh *et al.*, 1994, 1995). It is discretized with $25m$ cubic cells, representing an uniform mesh of $801 \times 801 \times 187$ nodes. The minimum and maximum velocities in the model are 2.2 km/s and 6.0 km/s respectively.

We resampled the grid with a grid interval of 75 m , which is consistent with a frequency of 7 Hz and a discretization rule of four grid points per minimum wavelength. The size of the resampled finite-difference grid is $268 \times 268 \times 64$. Ten points PML layers were added along each side of the FD grid for infinite model, noted test 1. The computing statistics are shown in table1.7. For test 2, we implemented a free surface at the top of the model. The computing

Table 1.7: Computing statistic of simulation in the Overthrust models: N_x, N_y and N_z denote the three dimensions of the model; Δt denotes the time step; The grid interval Δh is idential in the three directions; T (s) denotes the time length of the simulation; The computed time of CPU is denoted by T_{cpu}; F (Hz) denotes the frequency extracted; P/λ denotes the spatial discretization, points per wavelength.

Model	N_x	N_y	N_z	Δt	Δh (m)	$T_{CPU}(s)$	T (s)	F (Hz)	P/λ
Overthrust infinite	288	288	84	0.005	75	466.9293	8	7	4.27
Overthrust free surface	288	288	74	0.005	75	396.2754	8	7	4.27

statistics are shown in table 1.7. The source is located at $x = 1.5\ km, y = 1.5\ km, z = 0.75\ km$, receivers are located at a horizontal plan at the depth of 10 m. As the receivers are not coincident with the grid nodes, the *Hicks* interpolation is used for an accurate solution. We perform the simulation using **TDM** approach with a time length of 8 s, and extracted the monochromatic wavefield at 7 Hz for comparison with the **DSM** approach. We obtain a very good agreement between the **TDM** (black) and **DSM** (gray) solutions, are shown in the figure 1.29. The test 2 is more realistic with a free surface on top of the model. Both source and receivers are close to the free surface. The results are shown in figure 1.30. We obtain a very good agreement between the two solutions as the infinite case.

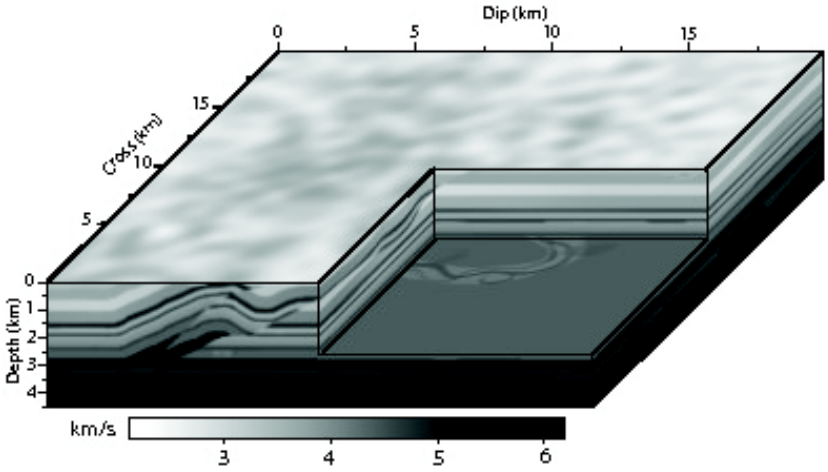

Figure 1.28: 3D view of the Overthrust V_P model.

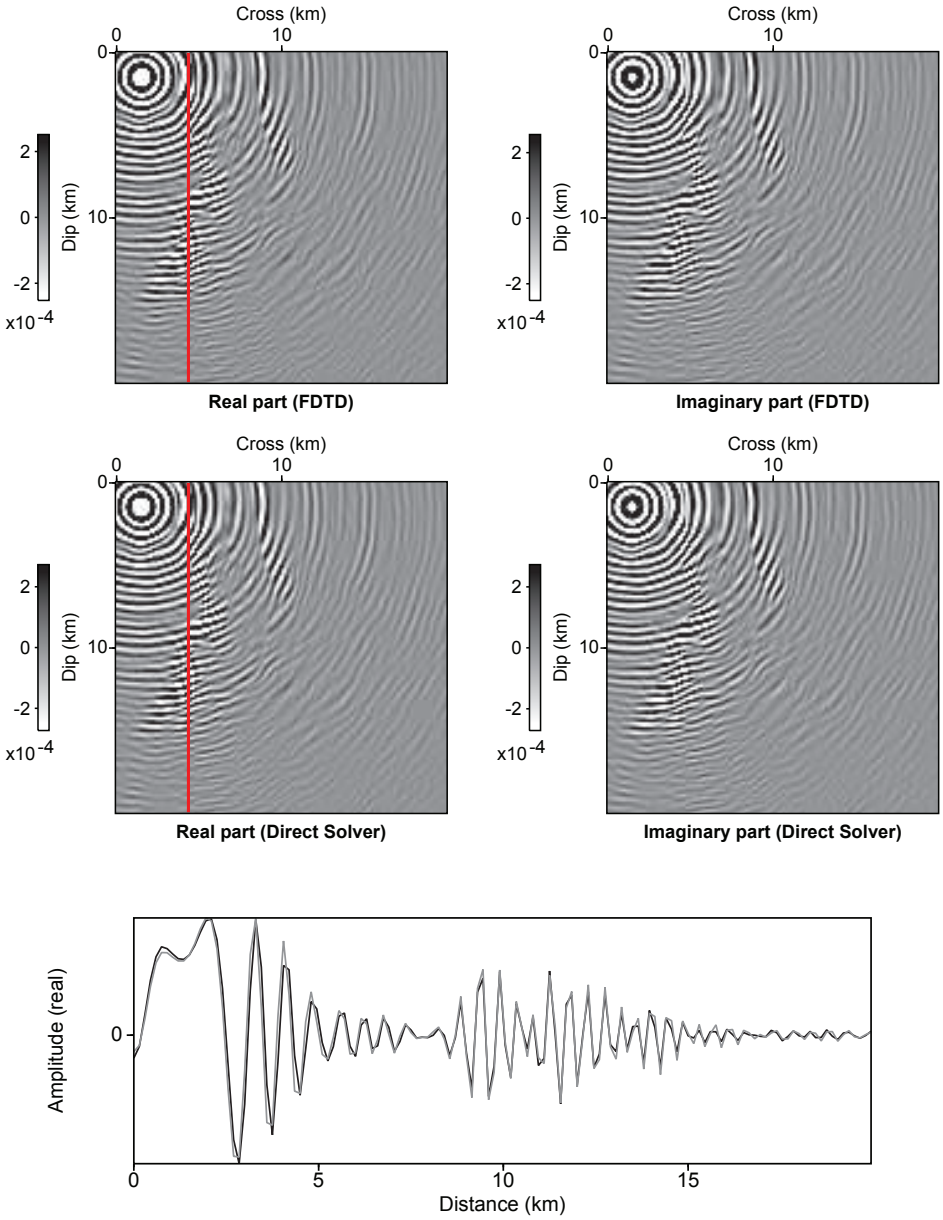

Figure 1.29: Comparison between monochromatic pressure wavefield computed with TDM and DSM approaches in the infinite space of Overthrust V_P model at 7 Hz. Top, the frequency domain solution computed with TDM combined with DFT, middle the solution calculated with DSM, bottom the comparison between the two solution extracted from the real part (red line), the black and gray lines are the TDM and DSM solutions, respectively. We observed a very good agreement between the two approaches.

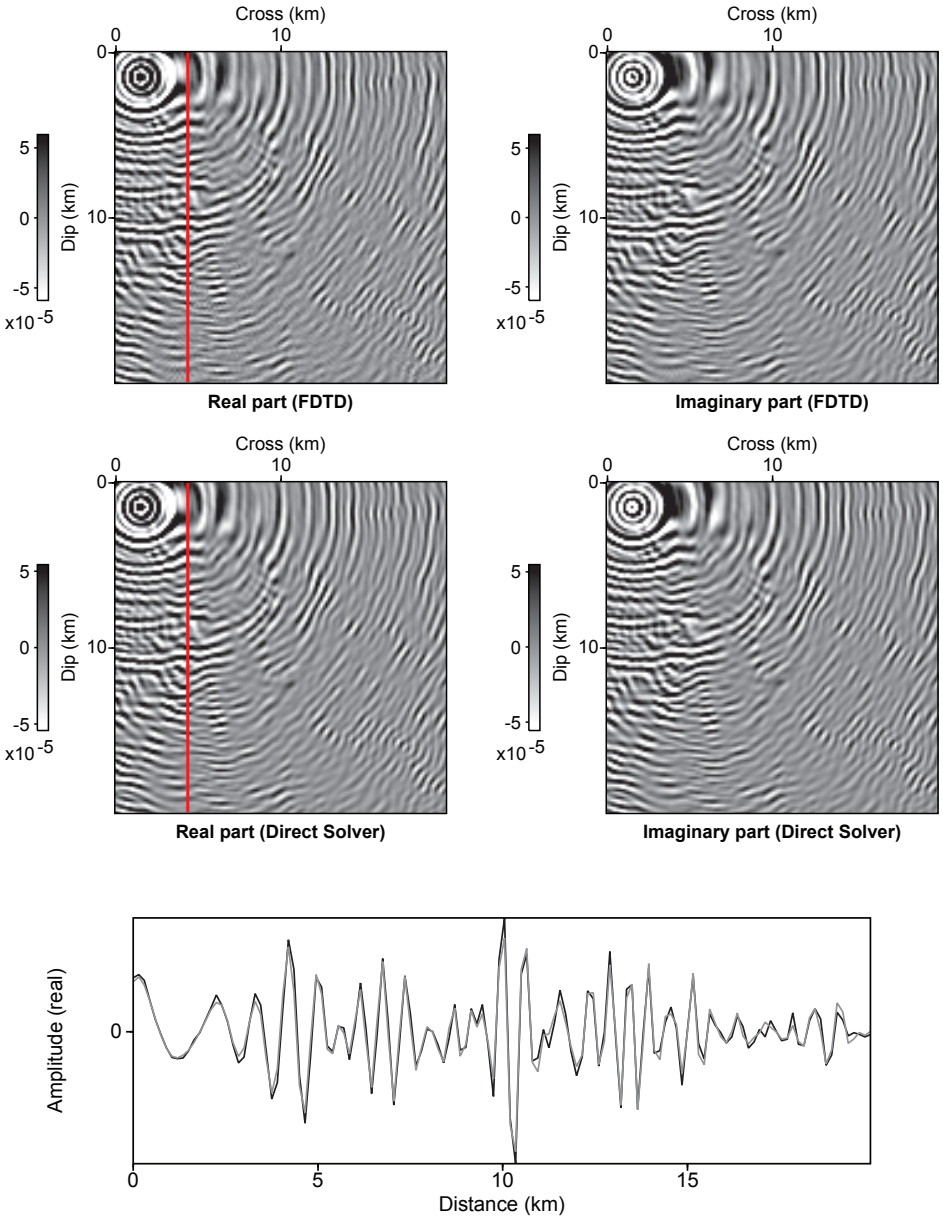

Figure 1.30: Comparison between monochromatic pressure wavefield computed with TDM and DSM approaches in the half space of Overthrust V_P model at 7 Hz. Top, the frequency domain solution computed with TDM combined with DFT, middle the solution calculated with DSM, bottom the comparison between the two solution extracted from the real part (red line), the black and gray lines are the TDM and DSM solutions, respectively. We observed a very good agreement between the two approaches.

Table 1.8: Computing statistic of simulation in the Overthrust models: N_x, N_y and N_z denote the three dimensions of the model; Δt denotes the time step; The grid interval Δh is idential in the three directions; T (s) denotes the time length of the simulation; The computed time of CPU is denoted by T_{cpu}; F (Hz) denotes the frequency extracted; P/λ denotes the spatial discretization, points per wavelength.

Model	N_x	N_y	N_z	Δt	Δ h (m)	$T_{CPU}(s)$	T (s)	F (Hz)	P/λ
SALT free surface	359	359	116	0.004	40	911.5153	8	7	5.35

1.6.2 The EAGE/SEG Salt model

The 3D SEG/EAGE salt model is a constant density acoustic model covering an area of 13.5 $km \times 13.5$ $km \times 4.2$ km (Aminzadeh *et al.*, 1997). The model is representative of Gulf Coast salt structure which contains salt sill, different faults, sand bodies and lenses (Figure 1.31). It is discretized with $20m$ cubic cells, representing an uniform mesh of $676 \times 676 \times 210$ nodes. The minimum and maximum velocities in the model are 1.5 km/s and 4.5 km/s. respectively.

We resampled the grid with a grid interval of 50 m to satisfy the discretization rule of 4 grid points per wavelength for a frequency of 7 Hz. The size of the resampled FD grid is $339 \times 339 \times 96$. Ten points PML layers are added for all of the sides of the model except the top, which represents the free surface. The computing statistics are shown in table 1.8. The source is located at $x = 1.56$ $km, y = 1.52$ $km, z = 0.01$ km, receivers are located at a horizontal plan at a depth of 5 m. We perform the simulation using **TDM** approach with the time length of 8 s, and extracted the monochromatic wavefield at 7 Hz to compare with the **DSM** approach. We obtain a very good agreement between the **TDM** (black) and **DSM** (gray) solutions, shown in the figure 1.32.

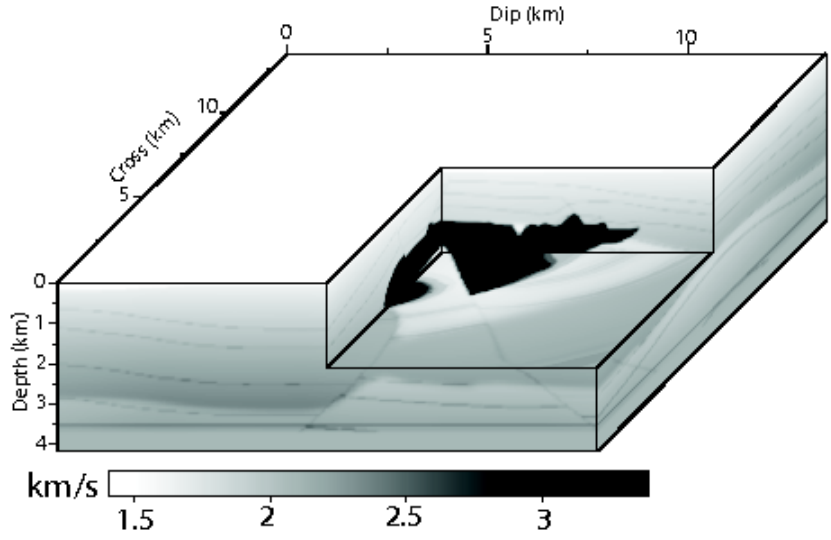

Figure 1.31: 3D view of the Salt V_P model.

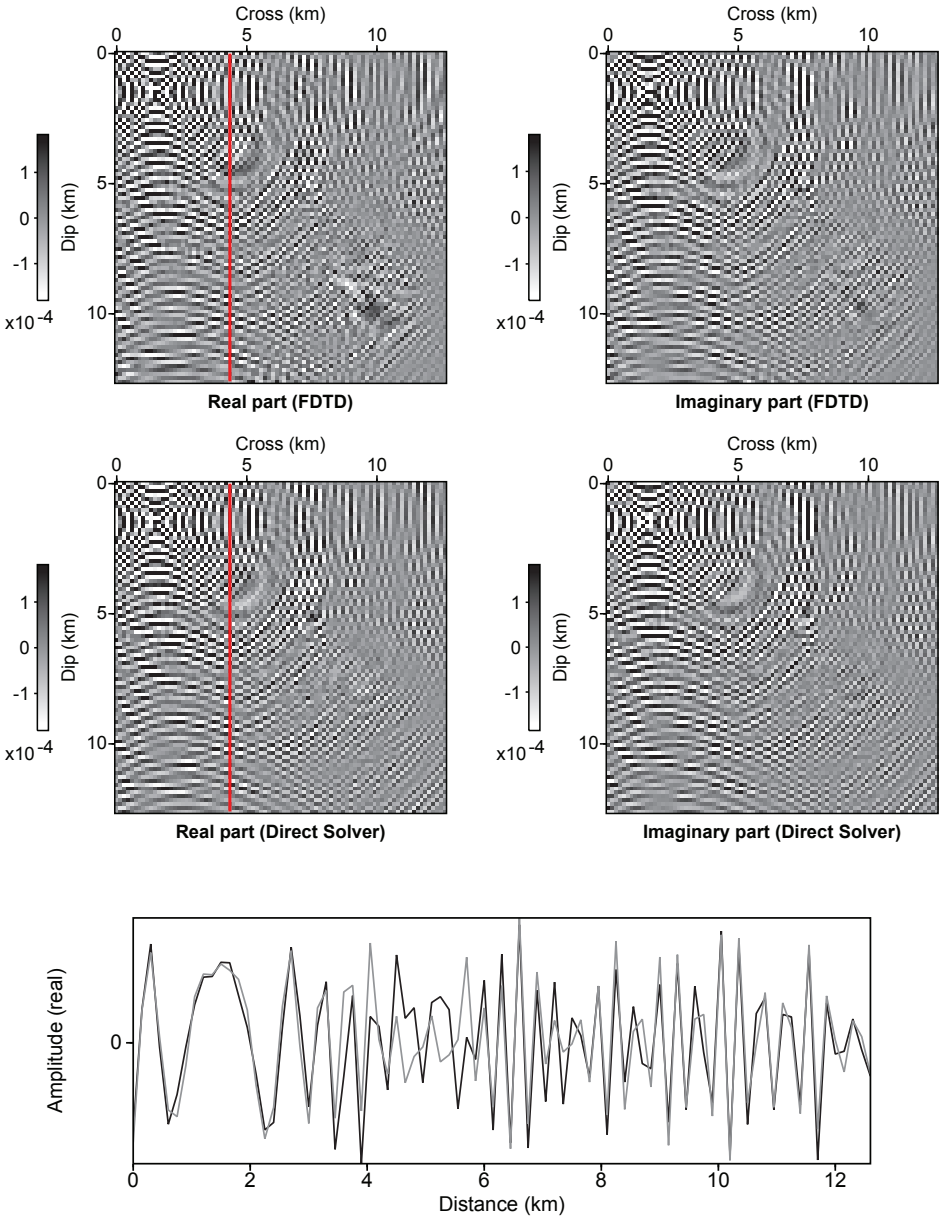

Figure 1.32: Comparison between monochromatic pressure wavefield computed with TDM and DSM approaches in the half space of Salt V_P model at 7 Hz. Top, the frequency domain solution computed with TDM combined with DFT, middle the solution calculated with DSM, bottom the comparison between the two solution extracted from the real part (red line), the black and gray lines are the TDM and DSM solution, respectively. We observed a very good agreement between the two approaches.

1.7 Partial conclusion for forward modeling

In this chapter, I have introduced the wave propagation modeling strategy in frequency and time domains. I have compared these two approaches, which have shown a very good agreement both for analytical and complex realistic models as those proposed by the industry (EAGE/SEG Overthrust and Salt models). Because of the memory requirement for the frequency approach, the choice of our strategy for solving the 3D acoustic forward problem is the time-domain staggered grid $\mathcal{O}(\Delta x^4, \Delta t^2)$ method. This method will be inserted into our imaging strategy for solving both the forward problem and the adjoint problem associated with the full waveform inversion.

I have investigated both the stability condition for the time integration and the optimal 4 grid-points per wavelength discretisation in order to avoid dispersion. I have extracted monochromatic solutions for frequencies embedded into the spectrum of the source used for time-domain simulation.

For considering the real source or receiver positions, we have estimated wavefields using an interpolation based on sinc function (*Hicks*). The *Hicks* interpolation allows us to implement the source at arbitrary position in coarse Cartesian grids. In particular, it allows for the implementation of the source nearby the free surface, an important issue, while constructing the solution at the real position of the source and the receiver. This is particularly important nearby the free surface. We have observed a good precision on the constructed numerical solutions.

The memory complexity is mitigated by the domain decomposition method as the processor is only in charge of a much smaller subdomain. However, due to the limit of subdomain and multi-sources used in the seismic prospecting, the parallelism on the sources is applied. Combining the two parallelizations if we have sufficient processor can significantly accelerate the computation. The methods of the parallelization make 3D FWI feasible besides of the development of the high performance computing.

The time-domain solver is a competitive approach thanks to its high degree of adaptability, simplicity of the implementation, and good scalability and robustness.

Part II

Inverse problem

In this part, I will discuss the full waveform inversion (FWI) in the frequency domain. FWI is a challenging data-fitting procedure based on full-wavefield modeling to extract quantitative information from seismograms. It relies on a formalism that allows to take into account the full information content of the data (Tarantola, 1984a) as opposed to more classical techniques such as travel time tomography. FWI was originally developed in the time domain (Tarantola, 1984a) whereas the frequency domain approach was proposed mainly in the 1990s by Pratt (1990a). The frequency domain formulation of FWI has been shown to be effective to build accurate velocity models for complex structures from global offset acquisition geometries (Pratt, 2004). As a result, FWI is a high resolution imaging process able to reach a spatial accuracy of half a wavelength (Sirgue et Pratt, 2004).

Issues of FWI are an efficient forward-modeling engine and a local optimization differential approach, in which the gradient and the Hessian operators are efficiently estimated. The gradient of the misfit function can be computed efficiently with the adjoint-state method (Plessix, 2006; Castellanos *et al.*, 2011). FWI is an ill-posed problem, that requires the starting model to be close enough to the real one in order to converge to the global minimum.

Another counterpart of FWI is the required computational resources when considering models and frequencies of interest. The task becomes even more challenging when one attempts to perform the inversion using the elastic equation (Shi *et al.*, 2007; Brossier *et al.*, 2009) instead of using the acoustic approximation (Mulder et Plessix, 2008; Barnes et Charara, 2009). This is the reason why, until recently, most studies were limited to 2D cases (e.g., Ravaut *et al.*, 2004). Recent advances in high-performance computing and multifold/multicomponent wide-aperture and wide-azimuth acquisitions make $3D$ acoustic FWI feasible today. FWI has focused a lot of interests and continuous efforts towards inversion of 3D data sets. Remarkable applications have been done in 3D using the acoustic approximation (Plessix, 2009; Sirgue *et al.*, 2010; Plessix et Perkins, 2010) but the extension to the 3D elastic case is still an on going work.

In this section, I shall focus on acoustic FWI performed in the frequency domain. The potential interest of such approach is to exploit the broad aperture coverage spanned by global-offset geometries to image a broad and continuous range of wavelengths in the velocity model. The frequency domain approach has been shown to be efficient for several reasons: First, only a few discrete frequencies are necessary to develop a reliable image of the medium, and second, proceeding sequentially from low to high frequencies defines a multi-resolution imaging strategy that helps to mitigate the non-linearity of the inverse problem (Pratt et Sams, 1996; Pratt *et al.*, 1998).

Firstly, I shall review the frequency domain FWI formulated as a least squares minimization problem which tries to minimize the misfit between the recorded and predicted seismic data. Then I shall present the different line search algorithms: Steepest-descent, Conjugate gradient, Gauss-Newton and Newton methods.

Secondly, I shall present a resolution analysis of the linearized inverse problem. We will discuss the adjoint-state method for computing the gradient of the misfit function (Plessix, 2006; Castellanos *et al.*, 2011), and its validation against finite difference. I shall introduce the

roles of the gradient and Hessian operators.

Finally, I shall discuss several validation tests. I shall show the efficiency of our algorithm with two levels of parallelism, and I shall discuss the strategy for 3D application. These learning will drive us towards real data applications as the Valhall case I shall consider in the objective of this work.

Chapter 2

Frequency domain waveform inversion

Contents

2.1 Theory

Full waveform inversion (FWI) aims to recover a quantitative representation of a subsurface model, which is parameterized by one or several physical properties that govern the propagation of seismic waves. The most usual parameter is the P-wave velocity varying with depth although

other quantities may be accounted for (shear-wave velocity, density, quality factor...). In seismic exploration, the data we consider are the seismic traces (i.e., seismograms) recorded at the surface by receivers. These receivers record the signal $\mathbf{d_{obs}}$ after it has propagated in the subsurface from the sources. The FWI is a technique that uses the information recorded by the receivers to recover the physical properties of the subsurface through the resolution of a numerical optimization problem. The concept of forward and inverse problem in the context of waveform inversion is schematically illustrated in figure 2.1. The non linear relationship, which

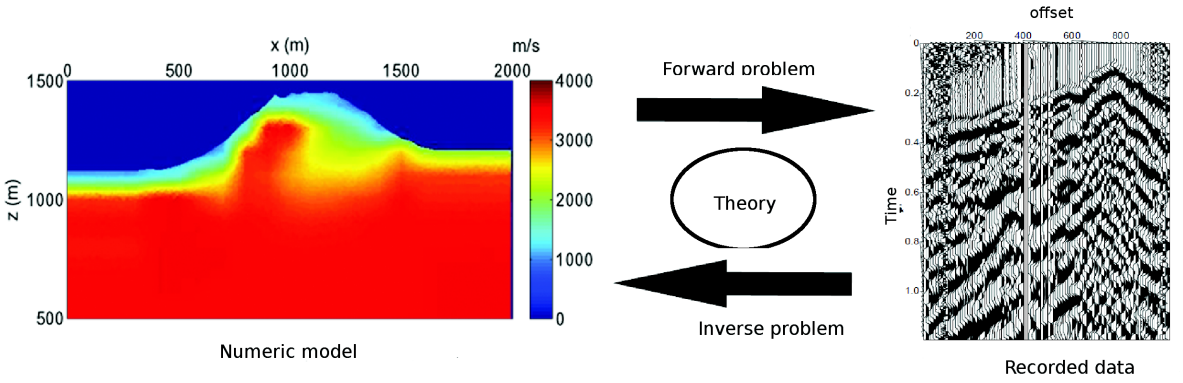

Figure 2.1: Forward and inverse problems in full waveform inversion. The inverse problem seeks to estimate a heterogeneous velocity model of the subsurface from seismic data recorded on the field.

relates the data $\mathbf{d}$ to the model $\mathbf{m}$, can be written in compact form as

$$\mathbf{d} = g(\mathbf{m}), \tag{2.1}$$

where $g(\mathbf{m})$ is a set of mathematical equation dependent on $\mathbf{m}$. A formal strategy to solve the inverse problem is to define the inverse operator g^{-1} such as the estimated solution is

$$\mathbf{m} = g^{-1}(\mathbf{d}). \tag{2.2}$$

In case this operator can be reconstructed, the imaging methods are referred to as direct methods. The resolution of the inverse problem using direct methods implies that the inverse of the operator can be found or at least estimated.

To avoid the estimation of the inverse operator g^{-1}, which could be a difficult and unstable task, another family of methods called local methods search incremental updates of the model. They estimate the model perturbation $\mathbf{\Delta m}$ of a starting model $\mathbf{m_0}$ such that the updated model is given by:

$$\mathbf{m} = \mathbf{m}_0 + \Delta\mathbf{m}. \tag{2.3}$$

The perturbation model $\mathbf{\Delta m}$ is found by minimizing a misfit between the observed data $\mathbf{d_{obs}}$ and forward modeled data $\mathbf{d_{cal}}(\mathbf{m})$ iteratively. These methods are local in the sense that the path leading to the solution depends on the initial guess $\mathbf{m_0}$ of the model $\mathbf{m}$. We define the

misfit vector $\mathbf{\Delta d} = \mathbf{d_{obs}} - \mathbf{d_{cal}}(\mathbf{m})$ of dimension N by the differences at the receiver positions between the observed data $\mathbf{d_{obs}}$ and the modeled data $\mathbf{d_{cal}}(\mathbf{m})$ for each source and each receiver of the seismic survey. Here $\mathbf{d_{cal}}$ is related to the modeled seismic wavefield $\mathbf{u}$ by a detection operator $\mathbf{R}$, which extracts the values of the wavefields computed in the full computational domain at the receiver positions for each source: $\mathbf{d_{cal}} = \mathbf{Ru}$. The updated model $\mathbf{m}$, as well as the initial model $\mathbf{m}_0$, represents some physical parameters of the subsurface discretized over the computational domain.

We will use the following notation for the forward problem (see Part I)

$$\mathcal{B}\left(\mathbf{m}(\mathbf{x}), \omega\right) \mathbf{u}\left(\omega, \mathbf{x}\right) = \mathbf{s}\left(\omega, \mathbf{x}\right), \tag{2.4}$$

where $\mathcal{B}$ is the forward modeling operator which relates linearly the wavefield $\mathbf{u}$ to the source $\mathbf{s}$ and non linearly the wavefield $\mathbf{u}$ to the model $\mathbf{m}$.

2.1.1 The linearization of the inverse problem

The local optimization problem is set up as a least-squares problem (Tarantola, 1987). The upcoming mathematical developments are independent of the domain in which we represent the data: time domain (Tarantola, 1984a,b; Lailly, 1984), frequency domain (Pratt et Worthington, 1990; Pratt, 1990b), Laplace domain (Shin et Cha, 2008), and so on.

The unconstrained/unweighted least-squares misfit function $\mathcal{C}(\mathbf{m})$ reads

$$\mathcal{C}(\mathbf{m}) = \frac{1}{2}\mathcal{L}^2(\Delta\mathbf{d}) = \frac{1}{2}\Delta\mathbf{d}^\dagger\Delta\mathbf{d}, \tag{2.5}$$

where † denotes the adjoint operator (transpose and conjugate operators denoted as ^{t} and *, respectively).

In the time domain, the implicit summation in the equation (2.5) is performed over the number of source-channel pairs and the number of time samples in the seismograms, where a channel is one component of a multi-component sensor. In the frequency domain, the summation over frequencies replaces the one over time. In the time domain, this vector is real-valued and, in frequency domain, it is complex-valued. Local optimization methods aims to find a minimum or a maximum of the misfit function in the vicinity of a starting model $\mathbf{m}_0$. The first step is to linearize the misfit function around $\mathbf{m}_0$ up to the second order, assuming that the misfit function is locally quadratic. This linearization relies on the assumption that only small perturbations in terms of amplitude and size hold.

Expanding to the second order the Taylor series of the misfit function $\mathcal{C}(\mathbf{m})$ around $\mathbf{m_0}$ with a perturbation model $\mathbf{\Delta m}$ gives

$$C(\mathbf{m}_0 + \Delta\mathbf{m}) = C(\mathbf{m}_0) + \sum_{j=1}^{N_p} \frac{\partial C(\mathbf{m}_0)}{\partial m_j}\Delta m_j + \frac{1}{2}\sum_{j=1}^{N_p}\sum_{k=1}^{N_p} \frac{\partial^2 C(\mathbf{m}_0)}{\partial m_j \partial m_k}\Delta m_j \Delta m_k + \mathcal{O}(\mathbf{m}^3), \tag{2.6}$$

where N_p denotes the dimension of the vector $\mathbf{m}$. The misfit function reaches its minimum when its first-order derivative vanishes. The first-order derivative with respect to a model parameter m_l reads:

$$\frac{\partial C(\mathbf{m})}{\partial m_l} = \frac{\partial C(\mathbf{m}_0)}{\partial m_l} + \sum_{j=1}^{N_p} \frac{\partial^2 C(\mathbf{m}_0)}{\partial m_j \partial m_l}\Delta m_j = 0, \tag{2.7}$$

and implies the compact expression with respect to $\mathbf{m}$,

$$\frac{\partial C(\mathbf{m})}{\partial \mathbf{m}} = \frac{\partial C(\mathbf{m_0})}{\partial \mathbf{m}} + \frac{\partial^2 C(\mathbf{m_0})}{\partial \mathbf{m}^2}\Delta\mathbf{m} = 0. \tag{2.8}$$

The gradient at the model $\mathbf{m_0}$ is denoted by $\mathcal{C}(\mathbf{m}_0)/\partial\mathbf{m}$, and the second derivative of the misfit function, $\partial^2\mathcal{C}(\mathbf{m}_0)/\partial\mathbf{m}^2$, is the Hessian. As $\partial\mathcal{C}(\mathbf{m})/\partial\mathbf{m}$ should be to zero at the minimum, the perturbation model vector is given by:

$$\Delta\mathbf{m} = -\left[\frac{\partial^2\mathcal{C}(\mathbf{m})}{\partial\mathbf{m}^2}\bigg|_{\mathbf{m}=\mathbf{m}_0}\right]^{-1}\frac{\partial\mathcal{C}(\mathbf{m})}{\partial\mathbf{m}}\bigg|_{\mathbf{m}=\mathbf{m}_0}. \tag{2.9}$$

2.1.2 The normal equations

The expression of the gradient as a function of the Fréchet derivatives is derived from the first order derivative of the misfit function with respect to m_l as

$$\begin{aligned}\frac{\partial C(\mathbf{m})}{\partial m_l} &= -\frac{1}{2}\sum_{i=1}^{N_d}\left[\frac{\partial d_{cal_i}}{\partial m_l}(d_{obs_i}-d_{cal_i})^* + (d_{obs_i}-d_{cal_i})\frac{\partial d^*_{cal_i}}{\partial m_l}\right]\\ &= -\sum_{i=1}^{N_d}\Re\left[\left(\frac{\partial d_{cal_i}}{\partial m_l}\right)^*(d_{obs_i}-d_{cal_i})\right],\end{aligned} \tag{2.10}$$

where N_d represents the number of terms in the data vector and is directly related to the acquisition, i.e., number of shots and receivers. The notation $\Re$ defines the real part of a complex number, while the expression * denotes the conjugate of a complex number. In matrix form, the equation (2.10) translates to

$$\begin{aligned}\nabla C_{\mathbf{m}} = \frac{\partial C(\mathbf{m})}{\partial \mathbf{m}} &= -\Re\left[\left(\frac{\partial \mathbf{d}_{cal}(\mathbf{m})}{\partial \mathbf{m}}\right)^t(\mathbf{d}_{obs}-\mathbf{d}_{cal}(\mathbf{m}))^*\right]\\ &= -\Re\left[\left(\frac{\partial \mathbf{d}_{cal}(\mathbf{m})}{\partial \mathbf{m}}\right)^\dagger(\mathbf{d}_{obs}-\mathbf{d}_{cal}(\mathbf{m}))\right] = -\Re\left[\mathbf{J}^\dagger\Delta\mathbf{d}\right],\end{aligned} \tag{2.11}$$

where $\mathbf{J}$ is the Fréchet derivatives or the sensitivity matrix.

Differentiation of the gradient expression (2.10), with respect to the model parameters gives the following expression in matrix form for the Hessian operator (Pratt *et al.*, 1998)

$$\frac{\partial^2 C(\mathbf{m})}{\partial\mathbf{m}^2} = \Re\left[\mathbf{J}^\dagger\mathbf{J}\right] + \Re\left[\frac{\partial\mathbf{J}^t}{\partial\mathbf{m}^t}(\Delta\mathbf{d}^*...\Delta\mathbf{d}^*)\right]. \tag{2.12}$$

Inserting the expression of the gradient operator (2.11) and the Hessian operator (2.12) into equation (2.9) gives for the following expression of the model perturbation,

$$\Delta\mathbf{m} = -\left\{\Re\left[\mathbf{J}^\dagger\mathbf{J} + \frac{\partial\mathbf{J}^t}{\partial\mathbf{m}^t}(\Delta\mathbf{d}^*...\Delta\mathbf{d}^*)\right]\right\}^{-1}\Re\left[\mathbf{J}^\dagger\Delta\mathbf{d}\right]. \tag{2.13}$$

Computation of the model perturbation requires the estimation of the gradient of the misfit function, which can be efficiently performed with the adjoint-state method (Plessix, 2006),

while the explicit estimation of the Hessian and its inverse is often not tractable from a computational viewpoint, because of the size of the data and model spaces. Some numerical methods exist to compute an approximation of the inverse of the Hessian recursively such as the L-BFGS method (Nocedal, 1980). Alternatively, second-order adjoint-state method allows one to compute efficiently Hessian-vector products without explicitly forming the Hessian. In this case, the normal-equation system can be solved with an iterative solver provided that an efficient preconditioner can be found (Métivier *et al.*, 2012). As a last resort, crude approximations of the Hessian such as diagonal approximation (Shin *et al.*, 2001b; Ravaut *et al.*, 2004) can be used to precondition the descent direction in steepest-descent or conjugate gradient algorithms, and hence speed up the convergence.

2.1.3 Computing the gradient with the adjoint-state method

The adjoint-state method (Lions, 1972; Chavent, 2009) is a well-known technique in inverse problem theory with many geophysical applications (Plessix, 2006; Castellanos *et al.*, 2011). Even though setting up the adjoint-state problem requires additional work as one has to find an adjoint variable, which has no primary physical interest in the solution of the inverse problem, this method is appealing, because the computation of the gradient with respect to a model parameter requires only two evaluations of the partial differential equations for each source. The alternative method, which consists in the explicit computation of the Fréchet derivatives, is expensive to compute, as it requires one forward modeling for each non-redundant position of sources and receivers (Shin *et al.*, 2001b) for getting the gradient.

Let us redefine the misfit function (2.5) as

$$\mathcal{C}(\mathbf{u},\mathbf{m}) = \frac{1}{2}\langle \mathbf{Ru}-\mathbf{d}_{obs}|\mathbf{Ru}-\mathbf{d}_{obs}\rangle. \tag{2.14}$$

We briefly review the application of the adjoint-state method in our implementation of FWI following the approach of Castellanos *et al.* (2011). Seismic modeling is performed through the resolution of the acoustic wave equation, which is formulated as a first-order velocity-pressure hyperbolic system in the time domain. The resulting forward-modeling operator is not self adjoint. The first step in the approach of Castellanos *et al.* (2011) is to transform the wave equation in pseudo-conservative form to make the forward modeling operator self adjoint. We use this equation as the state equation in the adjoint-state method. We will see that this approach allows us to use the same forward modeling engine to compute the state and the adjoint-state variables, and provides us a recipe to derive the correct source of the adjoint-state equation. Moreover, the scattering kernel of the gradient of the misfit function is made independent to the differential operator, and hence of the discretization method that is used to solve the forward problem. This provides a suitable framework to interface different modeling engines with the inversion.

The first-order acoustic system is written as (1.9). Transferring all the physical variables in the left-hand side leads to a pseudo-conservative form of the wave equation:

$$\begin{aligned}
\frac{1}{\kappa(x,y,z)}\frac{\partial p(x,y,z,t)}{\partial t} &= \frac{\partial v_x(x,y,z,t)}{\partial x}+\frac{\partial v_y(x,y,z,t)}{\partial y}+\frac{\partial v_z(x,y,z,t)}{\partial z}\\
\rho(x,y,z)\frac{\partial v_x(x,y,z,t)}{\partial t} &= \frac{\partial p(x,y,z,t)}{\partial x}
\end{aligned}$$

$$\begin{aligned}\rho(x,y,z)\frac{\partial v_y(x,y,z,t)}{\partial t} &= \frac{\partial p(x,y,z,t)}{\partial y}\\ \rho(x,y,z)\frac{\partial v_z(x,y,z,t)}{\partial t} &= \frac{\partial p(x,y,z,t)}{\partial z}.\end{aligned} \tag{2.15}$$

We write equation 2.15 in compact form as

$$\Lambda\partial_t\mathbf{u} - \mathcal{B}'\mathbf{u} = \mathbf{s}', \tag{2.16}$$

where Λ and $\mathcal{B}'$ are defined by

$$\Lambda = \begin{pmatrix} \frac{1}{\kappa} & 0 & 0 & 0\\ 0 & \rho & 0 & 0\\ 0 & 0 & \rho & 0\\ 0 & 0 & 0 & \rho \end{pmatrix}, \qquad \mathcal{B}' = \begin{pmatrix} 0 & \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z}\\ \frac{\partial}{\partial x} & 0 & 0 & 0\\ \frac{\partial}{\partial y} & 0 & 0 & 0\\ \frac{\partial}{\partial z} & 0 & 0 & 0 \end{pmatrix},$$

and $\mathbf{s}' = \Lambda\mathbf{s}$. Of note, the matrix $\mathcal{B}'$ is symmetric, because all the physical properties were moved on the left-hand side of equation 2.15.

We consider the augmented functional L subject to the constraint that the state equation is satisfied:

$$L(\mathbf{u},\mathbf{p}_1,\mathbf{m}) = J(\mathbf{u},\mathbf{m}) + \langle \mathbf{p}_1 \mid \left(\Lambda\partial_t\mathbf{u} - \mathcal{B}'\mathbf{u} - \mathbf{s}'\right)\rangle \tag{2.17}$$

where $\langle f(x,t)g(x,t)\rangle = \int\limits_{\Omega}\int_0^T f^*(x,t)g(x,t)dtdx$ and fields u, p_1, and m, are assumed independent. The adjoint-state variable is denoted by $\mathbf{p}_1$. The state equation associated with the adjoint variable $\mathbf{p}_1$ is the wave equation written in pseudo-conservative form.

Of note, we have $J(\mathbf{u},\mathbf{m}) = C(\mathbf{m})$, where $\mathbf{u}$ and $\mathbf{m}$ are processed as two independent variables, for any realization $\mathbf{u}$ of the state equation. The functional J is defined for time-domain data, as we perform seismic modeling in the time domain:

$$\begin{aligned}J(u,m) &= \frac{1}{2}\sum_r\int_0^T \left(\mathbf{R}\mathbf{u}(\mathbf{x},\mathbf{t}) - \mathbf{d}_{obs_r(t)}\right)^2 dt\\ &= \frac{1}{2}\sum_r\int_0^T\int_\Omega \left(\mathbf{u}(\mathbf{x},\mathbf{t}) - \mathbf{d}_{obs(x,t)}\right)^2 \delta(x-x_r)dxdt,\end{aligned} \tag{2.18}$$

where $\mathbf{R}$ is a sampling operator that extracts the values of the wavefield at the receiver positions. The adjoint field $\mathbf{p}_1$ can be referred to as Lagrange multiplier associated with the Lagrangian L. At the saddle points (u,p_1) of the Lagrangian L, we should have $\partial L/\partial p_1 = 0$, $\partial L/\partial u = 0$, which define the state and the so-called adjoint-state equations, respectively.

When these equations are satisfied, we have:

$$\frac{\partial\mathcal{C}}{\partial\mathbf{m}} = \frac{\partial L}{\partial\mathbf{m}}, \tag{2.19}$$

which provides to us the expression of the gradient we are interested in. Equation 2.19 provides an explicit expression of the gradient of $\mathcal{C}$ as a function of the state and adjoint-state variables.

The condition $\partial L/\partial p_1 = 0$ provides the state equation:

$$\Lambda \partial_t \mathbf{u} + \mathcal{B}'\mathbf{u} = \mathbf{s}'. \tag{2.20}$$

In practice, we compute numerically $\mathbf{u}$ by coming back to the non-conservative form of the wave equation:

$$\partial_t \mathbf{u} + \mathcal{B}\mathbf{u} = \mathbf{s}, \tag{2.21}$$

where it is reminded that $\mathbf{s}' = \Lambda \mathbf{s}$. The relation $\partial L/\partial u = 0$ requires a rewriting of the Lagrangian using derivation by parts which gives:

$$\begin{aligned} L(\mathbf{u}, \mathbf{p}_1, m) &= \mathcal{C}(\mathbf{u}, \mathbf{m}) + \int_\Omega \left(\mathbf{p}_1(T)\Lambda\mathbf{u}(T) - \mathbf{p}_1(0)\Lambda\mathbf{u}_0\right) dx \\ &- \langle \partial_t \mathbf{p}_1 | \Lambda \mathbf{u}\rangle - \langle \mathbf{p}_1 | \left(\mathcal{B}'\mathbf{u} + \mathbf{s}'(t)\right)\rangle, \end{aligned} \tag{2.22}$$

where T is the final time. Perturbing the variable $\mathbf{u}$ in the Lagrangian in direction $\mathbf{z}$, we can find the derivative as

$$\frac{\partial L}{\partial \mathbf{u}} \cdot \mathbf{z} = -\langle \Lambda^\dagger \partial_t \mathbf{p}_1 | \mathbf{z}\rangle - \langle \mathcal{B}'^\dagger \mathbf{p}_1 | \mathbf{z}\rangle + \int_\Omega \mathbf{p}_1(T)\mathbf{z}(T) dx, \tag{2.23}$$

whatever is the value of $\mathbf{z}$. If we impose this derivative to be zero, and if we consider the final condition $\mathbf{p}_1(T) = 0$, we have the adjoint-state equation:

$$\Lambda^\dagger \partial_t \mathbf{p}_1 + \mathcal{B}'^\dagger \mathbf{p}_1 = \mathbf{R}^t \left(\mathbf{R}\mathbf{u} - \mathbf{d}_{obs}\right). \tag{2.24}$$

As we inferred from the chapter 1.1.2, Λ is symmetric and real, and $\mathcal{B}'^\dagger = -\mathcal{B}'$. Therefore, the adjoint-state equation (2.24) can be simplified as:

$$\Lambda \partial_t \mathbf{p}_1 - \mathcal{B}'\mathbf{p}_1 = \mathbf{R}^t \left(\mathbf{R}\mathbf{u} - \mathbf{d}_{obs}\right). \tag{2.25}$$

It shows that the field $\mathbf{p}_1$ satisfies the conservative wave equation, where the source term is the data residuals $\mathbf{R}^t \left(\mathbf{R}\mathbf{u} - \mathbf{d}_{obs}\right)$.

As for the state equation, we can transform back this equation in non conservative form for numerical resolution:

$$\partial_t \mathbf{p}_1 - \mathcal{B}\mathbf{p}_1 = \Lambda^{-1}\mathbf{R}^t \left(\mathbf{R}\mathbf{u} - \mathbf{d}_{obs}\right). \tag{2.26}$$

Equation 2.26 shows that the adjoint-state variable can be computed with the non conservative form of the wave equation (i.e., the one implemented in the FWI code) provided that the source term is properly scaled by the coefficient of the diagonal matrix Λ. If receivers are hydrophones, the adjoint sources, which is implemented on the pressure component, must be scaled by the bulk modulus, while the velocity components of the adjoint source must be scaled by density if receivers are geophones.

The first order partial derivative of Lagrangian (2.17) with respect to the model parameter m_i leads to

$$L(\mathbf{m} + \epsilon\mathbf{z}) - L(\mathbf{m}) = \epsilon \langle \mathbf{p}_1(t) | \frac{\partial \Lambda}{\partial m_i} \mathbf{z} \left(\partial_t \mathbf{u} - \mathbf{s}\right)\rangle. \tag{2.27}$$

Dividing by the small parameter gives the expression of the gradient $\partial L/\partial m_i$. Using condition (2.19), we infer

$$\frac{\partial \mathcal{C}}{\partial m_i}(\mathbf{u}) = -\left(\partial_t \mathbf{u} - \mathbf{s}\right)^t \left(\frac{\partial \Lambda}{\partial m_i}\right)^T \mathbf{p}_1. \tag{2.28}$$

In frequency domain, it can be written as

$$\frac{\partial \mathcal{C}}{\partial m_i}(\mathbf{u}_m) = \Re\left(\sum_\omega \left(i\omega \mathbf{u} + \mathbf{s}\right)^\dagger \left(\frac{\partial \Lambda}{\partial m_i}\right)^\dagger \mathbf{p}_1\right). \tag{2.29}$$

It is worth noting that the scattering kernel of the gradient, $\frac{\partial \Lambda}{\partial m_i}$ in equation 2.29, is diagonal, which facilitates parallel implementation of the gradient based on domain decomposition, and, more importantly does not depend on differential operators, that makes the gradient building independent of the modeling engine used in the FWI code. We will use this property to interface different modeling engines based on different kinds of meshing, such as finite-difference methods on Cartesian grids and finite-element discontinuous Galerkin methods on unstructured tetrahedral meshes, with the inversion.

2.1.3.1 Interpretation of the gradient and resolution analysis

We can clearly see from equations (2.29) that the gradient is the result of the product between the incident field $\mathbf{u}$ and the back-propagated field $\mathbf{p}_1$. This implies that the gradient is computed in two steps : (1) compute the incident wavefield for the source $\mathbf{s}$; (2) compute the back-propagatied field by using the virtual residual source $\Lambda^{-1}(\mathbf{Ru} - \mathbf{d}_{obs})^*$, equation 2.26. If the background model is sufficiently accurate, the incident wavefield and the back-propagated adjoint wavefield should arrive at the same time at the position of the missing heterogeneity that generates the residual in the adjoint-wavefield source: the zero-lag correlation of the incident and adjoint wavefield provides the contribution of the source-receiver pair in the perturbation model at the position of the model parameter. Alternatively, the gradient can be interpreted as the zero-lag correlation of the partial derivative wavefield at the receiver positions with the data residuals: both signals can be interpreted as the wavefield scattered by the missing heterogeneities in the background model (Pratt *et al.*, 1998).

A resolution analysis of the gradient through plane-wave decomposition in homogeneous background media is possible (Sirgue et Pratt, 2004). The general theoretical framework is that of inverse scattering and diffraction tomography (Miller *et al.*, 1987; Wu et Toksöz, 1987). We consider a homogeneous medium as well as one source and one receiver with the velocity c_0 (Figure 2.2). The direction of the plane wave from the source and the receiver to the diffraction point is denoted by $\vec{\mathbf{s}}$ and $\vec{\mathbf{r}}$, respectively. By replacing the Green's functions with the plane-wave approximations, assuming that we are in the far field, and that the amplitude effects could be ignored, we have

$$\begin{aligned} G_0(\mathbf{x}, \mathbf{s}) &\approx exp(ik_0 \vec{\mathbf{s}}.\mathbf{x}) \\ G_0(\mathbf{x}, \mathbf{r}) &\approx exp(ik_0 \vec{\mathbf{r}}.\mathbf{x}), \end{aligned} \tag{2.30}$$

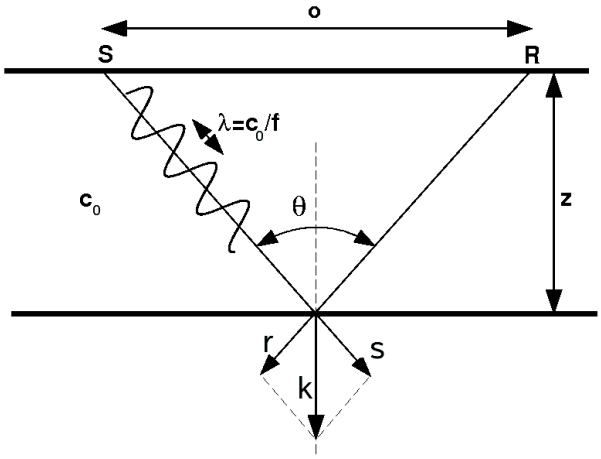

Figure 2.2: Wavenumber illumination. One source-receiver pair and one frequency in the data space maps one wavenumber in the model space. The local resolution at the diffraction point is controlled by the coverage of the wavenumber vector $\mathbf{k}$.

where $k_0 = \omega/c_0$ is the wavenumber of the incident wave. Inserting the equation (2.30) into (2.29), we have

$$\nabla \mathcal{C}_{\mathbf{m}} = -\omega^2 \Re \left\{ exp(-ik_0(\vec{\mathbf{s}} + \vec{\mathbf{r}}).\mathbf{x}) \Delta d \right\}. \tag{2.31}$$

This equation is, therefore, an inverse Fourier summation in which the weight function for each source/receiver pair is determined by the data residuals Δd (Sirgue et Pratt, 2004). The coverage of the wavenumber vector $k_0(\vec{\mathbf{s}} + \vec{\mathbf{r}})$ defines the spatial resolution with which the gradient is built. The ratio between the source-receiver offset and depth, that defines the scattering angle θ, the frequency, and the local wavespeed (these two last parameters defining the local wavelength) can be related to the following expression:

$$\mathbf{k} = \frac{2\omega}{c_0} \cos(\theta/2)\, \vec{\mathbf{n}}, \tag{2.32}$$

where $\vec{\mathbf{n}}$ is a unit vector in the direction of $\vec{\mathbf{s}} + \vec{\mathbf{r}}$. This relationship is illustrated schematically in figure 2.2 for a horizontal reflector at the base of a homogeneous half space. Of note, the relationship 2.32 highlights the redundancy of multi-fold seismic experiments. While the model space is two dimensional and can be parameterized by the two components of the wavenumber vectors, the data space is three dimensional, and can be parameterized by the angular frequency, and the source and receiver positions. Alternatively, the scattering and the incidence angles can replace the source and receiver positions in the data-space parameterization (Thierry *et al.*, 1999). The equation (2.32) tells us that, for surface acquisition, the optimum resolution for a given frequency is obtained at zero offset, and corresponds to half the wavelength, $\lambda/2$. Moreover, low frequencies and large offsets map small wavenumbers, giving information of the large-scale structures of the subsurface, while high frequencies and short offsets map the high wavenumbers of the subsurface. However, note that long offsets involve propagation of waves over larger distances, along which traveltime errors are accumulated. This makes the FWI of long-offset data to be more non linear, and hence subject to cycle skipping artifacts (Sirgue, 2006; Pratt, 2008; Virieux et Operto, 2009).

2.1.3.2 Numerical validation of the gradient

To validate the implementation of the gradient performed with the adjoint-state method, we compare the values of the gradient obtained with the adjoint-state method and with a finite-difference method. The finite-difference gradient approximation is given by

$$\frac{C(\mathbf{m}+\Delta\mathbf{m})-C(\mathbf{m})}{\Delta\mathbf{m}} \approx \frac{\partial C(\mathbf{m})}{\partial \mathbf{m}}. \tag{2.33}$$

We recall that C denotes the misfit function and $\mathbf{m}$ is the subsurface model.

We perform a validation test in an infinite homogeneous background with a sphere located in the middle of the mesh (r = 50 m). The computational grid is $41 \times 41 \times 41$ with 4 points PML in each direction. The grid interval is $\Delta h = 50\ m$. The temporal source excitation is a Ricker wavelet with a central frequency of 5 Hz. It is located in the center of the mesh at 100 m in depth. A perfect acquisition is used, which means that the receivers are uniformly deployed at each node of the computational grid. The velocity is 4000 m/s and 4300 m/s in the background model and in the sphere, respectively. The gradient is calculated in the framework of the acoustic approximation using a $\mathcal{O}(\Delta x^4, \Delta t^2)$ finite-difference modeling engine. The modeling and the inversion are performed in the time domain and in the frequency domain, respectively. The gradient is built for a single frequency of 5 Hz. Gradients computed with the adjoint-state method and the finite-difference method are shown in figures 2.3 and 2.4, respectively. Direct comparison between the two gradients is shown in figure 2.5. We show a good agreement between the two solutions. The difference in amplitude between the two solutions might result from the discretization (figure 2.5). A smaller step size is desired to reduce the truncation error $\mathcal{O}(m)$, but would also increase the numerical rounding errors. The error around 21% with a step of 300 m/s seems acceptable. Therefore, we conclude that the adjoint-state model is properly implemented.

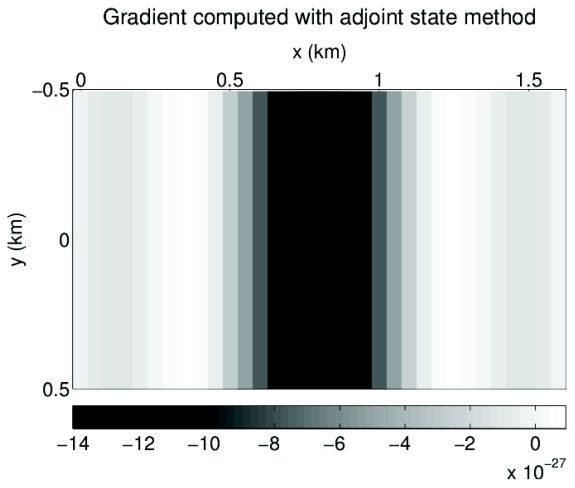

Figure 2.3: Computation of the gradient with the adjoint-state method. A horizontal plan at a depth of 800 m.

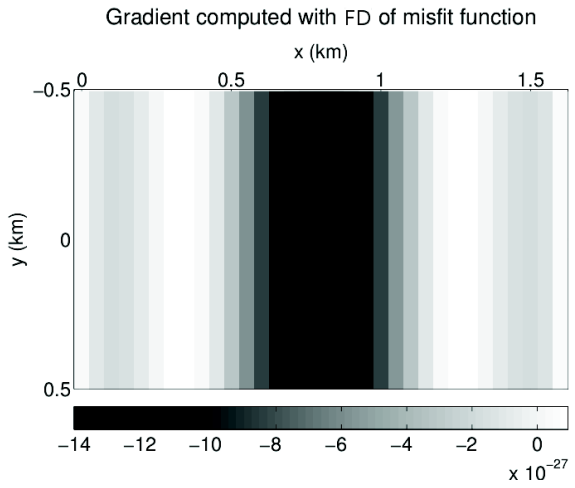

Figure 2.4: The same as the figure 2.3 but computed from the FD approach.

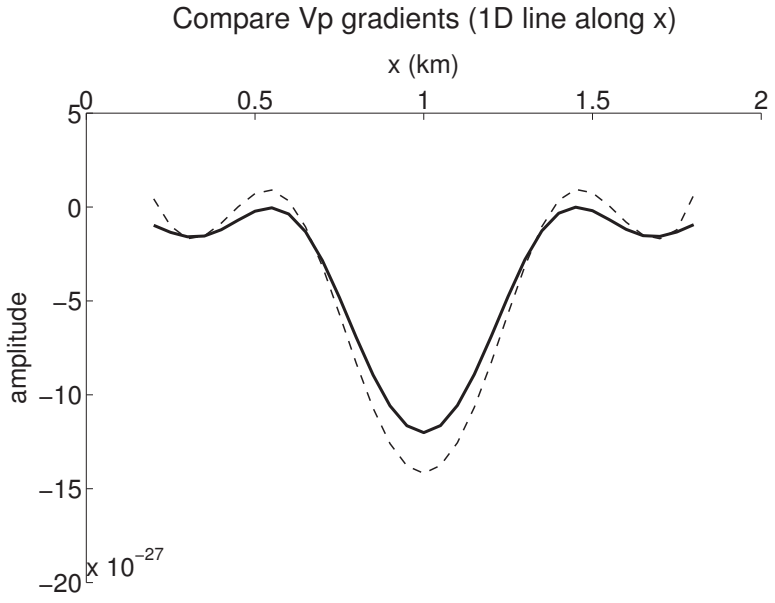

Figure 2.5: Comparison between profiles extracted from the gradients computed with the adjoint-state method (continuous black line) and the finite-difference method (dashed gray line). The profile is extracted at the position y = 800 m and z = 800 m. The error obtained is RMS = 0.213. Differences may come from the FD discretization.

2.1.4 Role of the Hessian

The first term of the Hessian, $\Re\left[\mathbf{J}^\dagger\mathbf{J}\right]$ (see equation (2.12)), is called approximate Hessian $\mathbf{H}_a$, which is formed by the zero-lag correlation between the partial-derivative wavefields with respect to different parameters. Therefore, it represents the spatial correlation between the images of different point scatterers. It can be viewed as a resolution operator resulting from limited bandwidth of the source and the acquisition geometry. Indeed, applying the inverse of the Hessian is equivalent to applying a spiking deconvolution to the gradient of the misfit function. The second term $\Re\left[\left(\frac{\partial\mathbf{J}_0}{\partial\mathbf{m}}\right)^T(\Delta\mathbf{d}_0^*\cdots\Delta\mathbf{d}_0^*)\right]$ is the zero-lag correlation between the second-order partial derivative of the wavefields with data residuals. Since first-order partial derivative is related to single scattering, it can be expected that second-order partial derivative is related to double scattering.

Due to the cost of the Hessian computation and storage, only manageable approximations are used. The size of the matrix $\mathbf{H}_a$ is $n_{nod}\times n_{nod}$. Each element in the matrix $\mathbf{H}_a$ is the scalar product of two partial derivative wavefields at the receiver positions, one of which is conjugated. This operation corresponds to zero-lag correlation in the time domain (Pratt *et al.*, 1998).

For one-dimensional problems, the $\mathbf{H}_a$ matrix is diagonally dominant and banded (in multi dimensions, the Hessian is a block matrix, where each block is diagonally dominant and banded), because the auto-correlation values are on the main diagonal and the frequency content is bandlimited. The gradient scaling performed by these diagonal terms allows to remove from the gradient the geometrical amplitude of the partial-derivative wavefields and the residuals. In the framework of surface seismic experiments, the scaling performed by the di-

agonal Hessian provides a good balance between shallow and deep perturbations (Pratt *et al.*, 1998; Virieux et Operto, 2009).

Computing explicitly the Hessian matrix for realistic problems is quite intensive and often impossible. Shin *et al.* (2001a) proposed a method for a fast but crude estimation of the Hessian, named Pseudo-Hessian $\mathbf{H}_p$. The expression is given by

$$\mathbf{H}_{p_{m_i,m_j}} = \Re\left[\mathbf{x}^\dagger \left[\frac{\partial \mathbf{B}}{\partial m_i}\right]^\dagger \left[\frac{\partial \mathbf{B}}{\partial m_j}\right]\mathbf{x}\right], \tag{2.34}$$

where $\mathbf{x}\,[\partial\mathbf{B}/\partial m_i]$ is the virtual-source matrix. This expression reduces the cost of the approximation of the Hessian to the cost of the gradient in terms of number of forward problems: computing the pseudo-Hessian requires two forward problems per source, instead of one forward problem per source and receiver. In practice, only the diagonal terms of the pseudo-Hessian is computed, because the computational effort required to compute the diagonal terms of the pseudo-Hessian is the same than the one required to compute the gradient.

2.1.5 Optimization algorithm: steepest-descent, conjugate gradient, quasi-Newton, Gauss-Newton and Newton algorithms

2.1.5.1 Steepest-descent method

The gradient of the misfit function represents the direction in which the misfit function is increasing most rapidly. Therefore, the misfit function can always be reduced by following the opposite of this direction. Pratt *et al.* (1998) gave a detailed interpretation of the gradient. The main idea of the steepest-descent method is to reduce the data residuals by iteratively updating the model according to

$$\mathbf{m}^{(k+1)} = \mathbf{m}^{(k)} + \alpha^{(k)} \nabla_m \mathcal{C}^{(k)}, \tag{2.35}$$

where k is an iteration number and α is a step length chosen to minimize the l_2 norm (2.5) in the opposite direction given by the gradient of $\mathbf{C}(\mathrm{m})$. Pratt *et al.* (1998) has given a simple formula to compute the step length α, as follow

$$\alpha^{(k)} = \frac{|\nabla_m \mathcal{C}|^2}{|\mathbf{J}\nabla_m \mathcal{C}|^2}, \tag{2.36}$$

where $|*|$ represents the Euclidean length of the vectors. For nonlinear problems, the step length must be found using line search techniques in the opposite direction of the gradient. In the framework of multi-parameter inversion, Sambridge *et al.* (1991) presented a subspace method, which allows to determine the step length for each parameter class.

Another linear search method is based on quadratic interpolation of the misfit function (Nocedal et Wright, 1999). For quadratic interpolation, we search two step lengths (α_1, α_2) such that:

$$\begin{aligned} \mathcal{C}(\mathbf{m}^{(k-1)} + \alpha_1 \Delta\mathbf{m}) &< \mathcal{C}(\mathbf{m}^{(k-1)}) \\ \mathcal{C}(\mathbf{m}^{(k-1)} + \alpha_2 \Delta\mathbf{m}) &> \mathcal{C}(\mathbf{m}^{(k-1)} + \alpha_1 \Delta\mathbf{m}), \end{aligned} \tag{2.37}$$

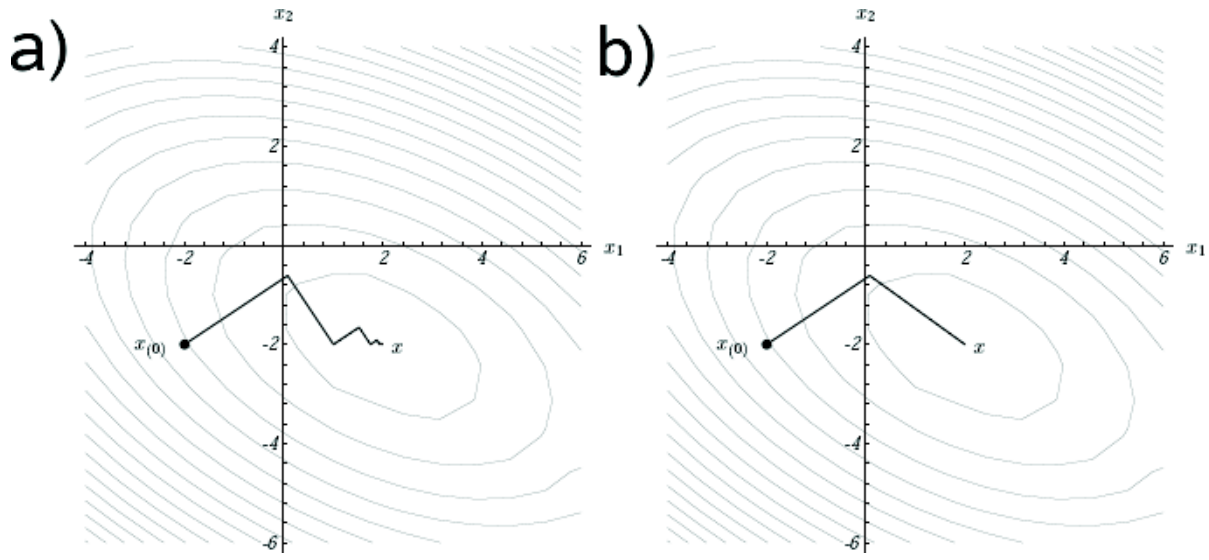

Figure 2.6: Sketch of optimization algorithms: (a) the steepest-descent method follows the path provided by the gradient of the misfit function; (b) the conjugate gradient combines two conjugate directions. This strategy accelerates the convergence (Brossier, 2009).

or

$$\begin{aligned} \mathcal{C}(\mathbf{m}^{(k-1)} + \alpha_1 \Delta\mathbf{m}) &> \mathcal{C}(\mathbf{m}^{(k-1)}) \\ \mathcal{C}(\mathbf{m}^{(k-1)} + \alpha_2 \Delta\mathbf{m}) &< \mathcal{C}(\mathbf{m}^{(k-1)} + \alpha_1 \Delta\mathbf{m}). \end{aligned} \tag{2.38}$$

The parabolic interpolation through the three points $(0, \mathcal{C}(\mathbf{m}^{(k-1)}))$, $(\alpha_1, \mathcal{C}(\mathbf{m}^{(k-1)} + \alpha_1 \Delta\mathbf{m}))$ and $(\alpha_2, \mathcal{C}(\mathbf{m}^{(k-1)} + \alpha_2 \Delta\mathbf{m}))$ allows us to find the minimum of the parabola $(\alpha_{min}, \mathcal{C}(\mathbf{m}^{(k)}))$ between 0 and α_2. The step length (α_{min} associated with this minimum is the step length that is used to update the model.

2.1.5.2 Conjugate gradient method

The conjugate gradient method is an effective method for solving a linear system. The governing idea is to move in non-interfering directions. The general form of the conjugate gradient direction is given by

$$\mathbf{p}^{(k)} = \mathcal{G}_{\mathbf{m}}^{(k-1)} + \beta^{(k)} \mathbf{p}^{(k-2)}, \tag{2.39}$$

where $\beta^{(k)}$ is a scalar that ensures that $p^{(k)}$ and $p^{(k-2)}$ are conjugate.

Figure 2.6a shows that the successive descent directions of the gradient method are orthogonal, resulting into a slow convergence. The conjugate gradient allows to combine the directions, usually, it can accelerate the convergence significantly (Figure 2.6b).

Fletcher et Reeves (1964) showed how to extend the linear conjugate gradient to nonlinear problems. Many variants of non-linear conjugate gradient methods were proposed since, the

most efficient one for waveform inversion being the Polak-Ribière method (Polak et Ribière, 1969). The coefficient $\beta^{(k)}$ is given by

$$\beta^{(k)} = \frac{\left(\mathcal{G}_{\mathbf{m}}^{(k-1)} - \mathcal{G}_{\mathbf{m}}^{(k-2)}\right)^t \mathcal{G}_{\mathbf{m}}^{(k-1)}}{\mathcal{G}_{\mathbf{m}}^{(k-2)^t} \mathcal{G}_{\mathbf{m}}^{(k-2)}}. \tag{2.40}$$

The conjugate gradient method remains the most popular optimization algorithm to perform FWI (Mora, 1987; Tarantola, 1987; Crase *et al.*, 1990).

2.1.5.3 Newton, Gauss-Newton, and quasi-Newton methods

The Newton direction is obtained by developing the misfit function (2.5) as a Taylor series and retaining terms up to quadratic order (Tarantola, 1987), which gives the expression

$$\mathcal{C}(\mathbf{p} + \delta\mathbf{p}) = \mathcal{C}(\mathbf{p}) + \delta\mathbf{p}^t \nabla_p \mathcal{C}(\mathbf{p}) + \frac{1}{2}\delta\mathbf{p}^t \mathbf{H} \delta\mathbf{p} + \mathcal{O}(\mid \delta\mathbf{p} \mid^3), \tag{2.41}$$

where $\mathbf{H}$ is the Hessian matrix.

The Newton method uses the full Hessian $\mathbf{H}$ (2.12). We look for the vector $\delta\mathbf{p}$ which is located at the minimum of the local quadratic approximation of the misfit function. For linear forward problems, this approach will converge in one iteration. The model perturbation computed with the Newton method is given by

$$\mathbf{H}\delta\mathbf{p} = -\nabla_p \mathcal{C} \qquad or \qquad \delta\mathbf{p} = -\mathbf{H}^{-1} \nabla_p \mathcal{C}. \tag{2.42}$$

As we may not need to know the Hessian with a great accuracy (Pratt *et al.*, 1998), the Gauss-Newton method only considers an approximate Hessian $\mathbf{H}_p$ (2.34) where the second-order scattering term in the Hessian is neglected. The perturbation model is given by

$$\delta\mathbf{p} = -\mathbf{H}_p^{-1} \nabla_p \mathcal{C}. \tag{2.43}$$

These methods are generally not used to solve large-scale inverse problems, because of the prohibitive cost of the explicit building of the Hessian. However, algorithms are emerging where efficient Hessian-vector product are computed with second-order adjoint-state method and the normal equations are solver with conjugate-gradient or Krylov iterative solvers (Epanomeritakis *et al.*, 2008; Métivier *et al.*, 2012). A key issue is to find a good preconditioner for the iterative solve of the normal-equation system, because two forward problems per source are required per iteration of the conjugate-gradient algorithm.

Quasi-Newton methods recursively estimate an approximation of the full Hessian. Hence, the Hessian is never built explicitly. The Quasi-Newton direction can be written as

$$\delta\mathbf{p} = -\mathbf{H}_q^{-1} \nabla_p \mathcal{C}. \tag{2.44}$$

where $\mathbf{H}_q$ is an approximation of the Hessian. The Hessian approximation $\mathbf{H}_q$ is chosen to satisfy

$$\mathbf{H}_q s_k = y_k, \tag{2.45}$$

where $s_k = (m_{k+1} - m_k)$ and $y_k = \nabla\mathcal{C}_{k+1} - \nabla\mathcal{C}_k$. This equation (2.45) is known as the "secant equation" (see (Nocedal et Wright, 1999) for details). The widespread expression of Quasi-Newton algorithm is the Broyden-Fletcher-Goldfarb-Shanno method (**BFGS**) (Nocedal et Wright, 1999) with its low-memory extension **L-BFGS** method (Nocedal, 1980) or the bounded one **L-BFGS-B** (Byrd *et al.*, 1995). The **BFGS** methods estimate the Hessian from the gradients and the solution vectors from the previous iterations, thereby increasing the Hessian accuracy with the iterations. The formula is given by

$$Hq_{k+1}^{-1} = \left(I - \frac{s_k y_k^T}{y_k^T s_k}\right) Hq_k^{-1} \left(I - \frac{y_k s_k^T}{y_k^T s_k}\right) + \frac{s_k s_k^T}{y_k^T s_k}, \tag{2.46}$$

where I is the identity matrix. **L-BFGS** algorithms keep in memory only limited number of vectors. The algorithm is built by two recursive loops, which allow to avoid the explicit computation of the Hessian as well as its storage. The product $\mathbf{H}_q^{-1}\nabla C$ is implicitly evaluated through inner products and vector summations involving ∇C and limited stored sequence $(s_i y_i)$ (Nocedal et Wright, 1999). Therefore **L-BFGS** is generally much faster than steepest-descent or conjugate gradient methods, and is quite competitive with Newton methods, in particular for weakly-contrasted media such as geological media where a limited amount of multiple scattering is recorded and guarantees that the Hessian remains positive-definite during its building (Métivier *et al.*, 2012).

2.1.6 Introducing regularization in FWI

FWI is a nonlinear ill-posed problem. Therefore, an infinite number of models may match the data, and the misfit function has many local minima. The inversion can be stuck into local minima, because of noise, incomplete representation of the wave physics, and of the experimental set-up (source directivity, receiver-ground coupling), inaccuracies of the initial model, and numerical errors (numerical dispersion, absorbing boundaries). To minimize the impact of these artifacts, some preconditioning and regularizations are conventionally applied to make the inversion better posed (Menke, 1984; Tarantola, 1987; Scales *et al.*, 1990). The misfit function (2.5) can be augmented as follows

$$\mathcal{C}(\mathbf{m}) = \frac{1}{2}\Delta\mathbf{d}^\dagger \mathbf{W}_d \Delta\mathbf{d} + \frac{1}{2}\sum_{i=1}^{N_p} \epsilon_i \left(\mathbf{m}_i - \mathbf{m}_{prior_i}\right)^\dagger \mathbf{W}_{m_i} \left(\mathbf{m}_i - \mathbf{m}_{prior_i}\right), \tag{2.47}$$

where $\mathbf{W}_d = \mathbf{S}_d^t \mathbf{S}_d$ is a preconditioning operator and $\mathbf{W}_{m_i} = \mathbf{S}_{m_i}^t \mathbf{S}_{m_i}$ are regularization operators associated with each parameter class $i = 1, ..., N_p$. The weighting operators $\mathbf{W}_d$ and $\mathbf{W}_m$ are the inverse of the data and model covariance operators in the framework of Bayesian formulation of FWI (Tarantola, 1987; Scales et Smith, 1994). The $\mathbf{S}_d$ corresponds to a weighting/filtering operator that is applied to the data. For example, it can be implemented as a diagonal weighting operator to weight data with respect to offset in order to strengthen the far-offset data in the case of crustal scale imaging (Operto *et al.*, 2006). The operator $\mathbf{S}_m$ is generally a roughness operator by taking a first or second-order finite difference operator. This implies that we apply smoothness constraint to the model parameters, as we minimize their roughness. This kind of regularization is generally called *Tikhonov regularizations* (Tikhonov, 1963; Tikhonov et Arsenin, 1977; Hansen, 1998). The hyperspace scalar parameters ϵ control the respective weight of the data-space and model-spaced misfit functions in equation (2.47).

Their values should be adapted to each parameter class. Minimization of the misfit function at the iteration k gives the Newton descent direction $\mathbf{p}_k$, written as

$$\mathbf{p}_k = -\left[\frac{\partial^2 C(\mathbf{m}_k)}{\partial \mathbf{m}^2}\right]^{-1} \frac{\partial C(\mathbf{m}_k)}{\partial \mathbf{m}}. \tag{2.48}$$

The update model is given by $\mathbf{m}_{k+1} = \mathbf{m}_k + \gamma_k \mathbf{p}_k$. The step length γ_k defines the amount of descent in the direction $\mathbf{p}_k$,

$$\begin{aligned}\mathbf{p}_k = \Re\left(\widehat{\mathbf{W}}_m^{-1}\mathbf{J}_k^{\dagger}\mathbf{W}_d\mathbf{J}_k + \widehat{\mathbf{W}}_m^{-1}\left(\frac{\partial \mathbf{J}_k^T}{\partial \mathbf{m}^T}\right)(\Delta\mathbf{d}_k^*...\Delta\mathbf{d}_k^*) + \Lambda\right)^{-1} \\ \Re\left(\widehat{\mathbf{W}}_m^{-1}\mathbf{J}_k^T\mathbf{W}_d\Delta\mathbf{d}_k^* + \Lambda\left(\mathbf{m}_k - \mathbf{m}_{prior}\right)\right),\end{aligned} \tag{2.49}$$

where Λ is a block diagonal damping matrix

$$\Lambda = \begin{pmatrix} \epsilon_1\mathbf{I}_M & ... & \mathbf{0} \\ ... & ... & ... \\ \mathbf{0} & ... & \epsilon_{N_p}\mathbf{I}_M \end{pmatrix}, \tag{2.50}$$

where $\mathbf{I}_M$ is the identity matrix of dimension M, M denotes the number of nodes in the computational mesh. The matrix $\widehat{\mathbf{W}}_m$ is a $N_p \times N_p$ block diagonal matrix, where each block is formed by the $\mathbf{W}_{m_i}$ matrices. The estimation of a step lenght γ_k is necessary, because of the locally-quadractic approximation of the misfit function. Of note, if we use $\mathbf{m}_{prior} = \mathbf{m}_k$, the regularization term in the gradient of the misfit function in equation (2.49) is dropped off.

2.1.7 On the choice of the norm in the data space

The least-squares norm is the most popular approach for frequency domain FWI (Pratt et Worthington, 1990; Pratt, 1990b). This norm requires the distribution of the misfit to be Gaussian (Tarantola, 1987). If this assumption is not satisfied, a poor result is expected. Therefore, the least-squares norm requires a careful quality control of the data to remove outliers. The $\mathcal{L}_2$ norm is usually written into the following form

$$\mathcal{C}_{\mathcal{L}_2}^{(k)} = \frac{1}{2}\Delta\mathbf{d}^{\dagger}\mathbf{S}_d^{\dagger}\mathbf{S}_d\Delta\mathbf{d}, \tag{2.51}$$

where $\Delta\mathbf{d} = \mathbf{d}_{obs} - \mathbf{d}_{cal}^{(k)}$. Synthetic data $\mathbf{d}_{cal}^{(k)}$ are computed at each iteration k by a forward-modeling engine defined by the operator $\mathbf{B}$. The gradient with respect to parameters can be written as

$$\mathcal{G}_{\mathcal{L}_2}^{(k)} = \Re\Big\{\mathbf{J}^t\mathbf{S}_d^{\dagger}\mathbf{S}_d\Delta\mathbf{d}^*\Big\}, \tag{2.52}$$

and the gradient can be computed with the adjoint-state method (Plessix, 2006), and is given by

$$\mathcal{G}_{m_i\,\mathcal{L}_2}^{(k)} = \mathcal{R}\Big\{\mathbf{u}^t\frac{\partial \mathbf{B}^t}{\partial m_i}\mathbf{B}^{-1^t}\mathbf{S}_d^{\dagger}\mathbf{S}_d\Delta\mathbf{d}^*\Big\}, \tag{2.53}$$

where the sparse matrix $\partial\mathbf{B}/\partial m_i$ is the radiation pattern of the diffraction by the model parameter m_i. As already mentioned, the data residuals are back-propagated in the medium

to compute the gradient of the misfit function when the least-squares norm is used in the data space. This implies that both the phase and the amplitude of the residuals contribute to the building of the gradient.

Unlikely the $\mathcal{L}_2$ norm, which theoretically suffers from poor robustness in the presence of large isolated and non-Gaussian errors, the least-absolute-value norm ($\mathcal{L}_1$) is less sensitive to noise in the framework of efficient frequency domain FWI (Brossier *et al.*, 2010c). Tarantola (1987); Crase *et al.* (1990) introduced the $\mathcal{L}_1$ norm in the time domain FWI, and Pyun *et al.* (2009) and Brossier *et al.* (2010c) used a $\mathcal{L}_1$ norm in the frequency domain FWI. The misfit function based on the $\mathcal{L}_1$ norm is given by

$$\mathcal{C}_{\mathcal{L}_1}^{(k)} = \sum_{i=1,N_d} |\mathbf{S}_{d_i}\Delta d_i|, \tag{2.54}$$

where $|\mathbf{S}_{d_i}\Delta d_i| = ((\mathbf{S}_{d_i}\Delta d_i)(\mathbf{S}_{d_i}\Delta d_i)^*)^{1/2}$, N_d is the number of elements in the misfit vector for one source and one frequency, and $\mathbf{S}_{d_i}$ are the elements of the diagonal $\mathbf{S}_d$. The gradient of the misfit function is given by

$$\mathcal{G}_{\mathcal{L}_1}^{(k)} = \Re\Big\{\mathbf{J}^t\mathbf{S}_d^t\mathbf{r}\Big\} = -\Re\Big\{\mathbf{u}^t\frac{\partial\mathbf{B}^T}{\partial m_i}\mathbf{B}^{-1^t}\mathbf{S}_d^\dagger\mathbf{r}\Big\} \qquad \text{with } r_i = \frac{\Delta d_i^*}{|\Delta d_i|} \quad \text{for } 1 \le i \le N, \tag{2.55}$$

where we assume that $|\Delta d_i| \neq 0$. In the case of real arithmetic numbers, the term $\Delta d_i^*/|\Delta d_i|$ corresponds to the *sign* function (Tarantola, 1987; Crase *et al.*, 1990).

For the $\mathcal{L}_1$ norm, the adjoint wavefield is computed by back-propagating the weighted data residuals that are normalized by their modulus (Brossier *et al.*, 2010c). The normalization of the residuals by their modulus gives some intuitive understanding why the $\mathcal{L}_1$ norm is less sensitive to large errors than the $\mathcal{L}_2$ norm. The $\mathcal{L}_2$ norm is chosen for this thesis work, because we shall use a high-quality three-dimensional data set, for which a careful quality control was performed.

2.1.8 Source estimation

In the case of real data application of FWI, the source term can be unknown, and hence needs to be estimated. In the frequency domain, the source signature can be estimated by solving a linear problem (Pratt, 1999) for each frequency. As the relationship between the seismic wavefield $\mathbf{p}(\omega)$ and the source $\mathbf{s}(\omega)$ is linear, we have

$$\mathbf{p}(\omega, \mathbf{x}) = \mathbf{s}(\omega)\mathbf{G}(\omega, \mathbf{x}), \tag{2.56}$$

where $\mathbf{G}$ is the Green function and $\mathbf{s}$ is the source wavelet. The misfit function (2.5) for the source estimation is

$$\begin{aligned} \mathcal{C}(\mathbf{m}) &= \frac{1}{2}(\mathbf{d}_o - \mathbf{d}_c)^T(\mathbf{d}_o - \mathbf{d}_c)^* \\ &= \frac{1}{2}(\mathbf{d}_o - \mathbf{s}\mathbf{G}_\mathcal{R})^T(\mathbf{d}_o - \mathbf{s}\mathbf{G}_\mathcal{R})^*, \end{aligned} \tag{2.57}$$

where $\mathbf{G}_\mathcal{R}$ is the restriction of Green function $\mathbf{G}$ at the receiver positions. The differentiation of (2.57) with respect to $\mathbf{s}$ leads to the updated expression of the source,

$$\mathbf{s}(\omega) = \frac{\mathbf{d}_o^T\mathbf{G}_\mathcal{R}^*}{\mathbf{G}_\mathcal{R}^T\mathbf{G}_\mathcal{R}^*}, \tag{2.58}$$

where it is reminded that $\mathbf{s}(\omega)$ is complex-valued.

The source estimation can be performed iteratively for each source and frequency during FWI, once the Green functions associated with the incident wavefields were computed. The source estimation is required to compute the data residuals and build the source of the adjoint-state equation. Of note, the source estimation proposed here relies on the assumption than the subsurface medium is known. This requires the alternating update of the subsurface medium and of the source in the iterative FWI algorithm. Generally, the source estimation is expected to become more accurate as the subsurface medium is updated by FWI. Therefore, the source estimation can also be used as a quality control of the FWI results (Brenders et Pratt, 2007; Malinowski *et al.*, 2011; Prieux *et al.*, 2011).

We validate our implementation of the source estimation in our FWI code with a synthetic experiment. The acquisition is composed of eight sources and 1178 receivers, which are uniformly distributed in an infinite homogeneous model. A Ricker wavelet with a central frequency of 5 Hz is used for computing the observed data. A Dirac wavelet is used as an initial guess of the source wavelet. We estimate the source signature for 16 frequencies between 0 Hz and 15 Hz. The frequency-domain data are extracted by discrete Fourier transform (DFT) from time-domain data. Figure 2.7 shows a comparison between the true and the reconstructed source wavelets. The reconstructed source wavelets were transformed back in the time domain by inverse Fourier transform from the 16 frequencies. We show a nearly perfect reconstruction. Then, we convolve the average of the reconstructed sources with the Green functions computed in the true model, and compare the observed data and the resulting data in frequency domain. Indeed, a very good agreement is shown between the two solutions for both the real (Figure 2.9) and imaginary (Figure 2.8) parts.

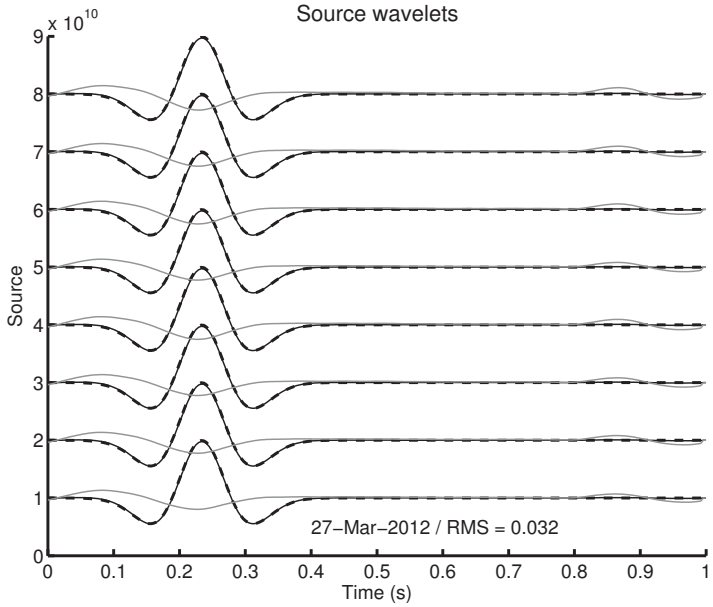

Figure 2.7: The source estimation: the original sources (black) and the reconstructed sources (dashed line). The gray line represent residuals.

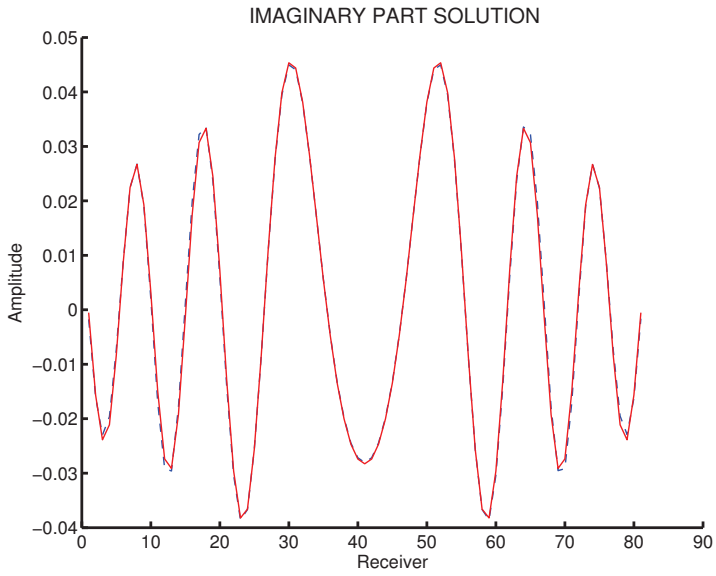

Figure 2.8: Comparison between frequency-domain data (imaginary part) computed with the true wavelet (continuous line) and data computed with the reconstructed wavelet (dashed line).

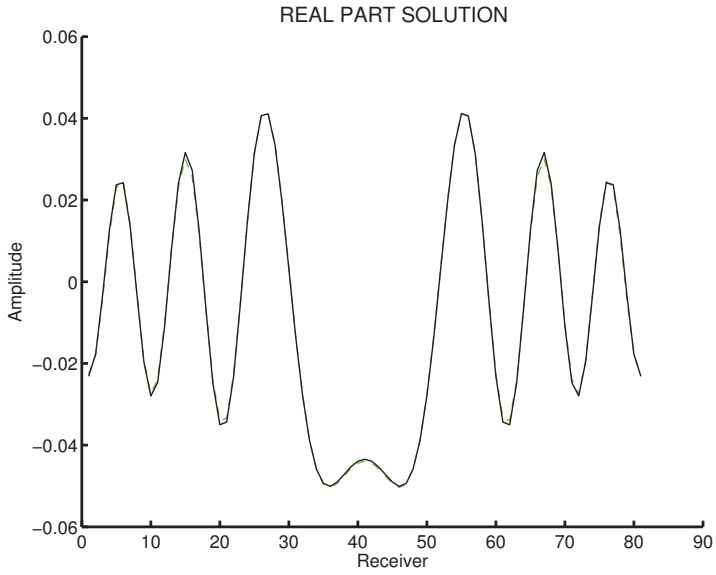

Figure 2.9: Same as figure 2.8 for the real part.

2.2 Algorithm

2.2.1 Multiscale approach of FWI

The full waveform inversion (FWI) has been developed in time and frequency domains. The first developments of FWI were performed in the time domain in the eighties (Tarantola, 1984a; Gauthier *et al.*, 1986; Mora, 1987). Pratt (1990b), and Pratt et Worthington (1990) proposed to recast FWI in the frequency domain in the nineties. Time-domain FWI allows one to select different wave packets in the seismograms, while frequency domain FWI promotes the decimation of the data in the frequency domain by selecting a few discrete frequencies to design computationally-efficient algorithms. A hierarchical approach, which proceeds from the low frequencies to the higher ones, is recommended to mitigate the non linearity of the FWI. This hierarchical approach can be implemented in the time domain (Bunks *et al.*, 1995) and in the frequency domain (Pratt, 1999), even if the latter provides the most natural framework for this.

As proposed by Bunks *et al.* (1995), we can progressively increase the high-frequency content of the data as the inversion progresses over iterations, while keeping involved in the inversion all of the previous frequencies. Alternatively, one may invert successively slightly overlapping frequency groups of progressively higher-frequency content to reduce the computational burden when modeling is performed in the frequency domain (in this case, the computational cost of modeling scales linearly with the number of frequencies) (Brossier *et al.*, 2009). The definition of the frequency bandwidth of each frequency groups should be driven by the best compromise between the need to avoid cycle-skipping artifacts resulting from the inversion of high frequencies, the need to simultaneously invert multiple frequencies to preserve a certain amount of data redundancy in the inversion, and the need to keep the computational cost reasonable by limiting the number of frequencies (Brossier *et al.*, 2009). Figure 2.11 illustrates two strategies to manage the frequencies in the full waveform inversion.

The frequency-domain provides a natural framework to implement multiscale FWI based on the hierarchical inversion of frequency groups of arbitrary bandwidth and sampling intervals (Virieux et Operto, 2009). Sirgue et Pratt (2004) proposed a rule for choosing the frequency interval: the governing idea is to reduce the wavenumber redundancy, which results from the double control of frequency and scattering angle on the wavenumber coverage. This redundancy can be reduced by limiting the inversion to a few discrete frequencies. The redundant control of frequency and scattering angle on the wavenumber coverage is shown by equation (2.32) and is illustrated in figure 2.2: one frequency and one scattering angle (or, aperture) in the data space map one wavenumber in the model space. This redundancy increases with the aperture bandwidth. Decimating the wavenumber redundancy allows one to limit the inversion to a few discrete frequencies. Therefore, it reduces the computational cost, and also allows for the management of a compact volume of data. The figure 2.11 from Sirgue et Pratt (2004) illustrates the rule for frequency selection: the highest wavenumber mapped by one frequency should be equal to the smallest wavenumber mapped by the next frequency. This leads to an increasing frequency interval with frequency.

In a homogenous medium and for a plane reflector, the guideline for frequencies selection is given by:

$$\Delta f_{n+1} = f_{n+1} - f_n = (1 - \alpha_{min}) f_{n+1}, \tag{2.59}$$

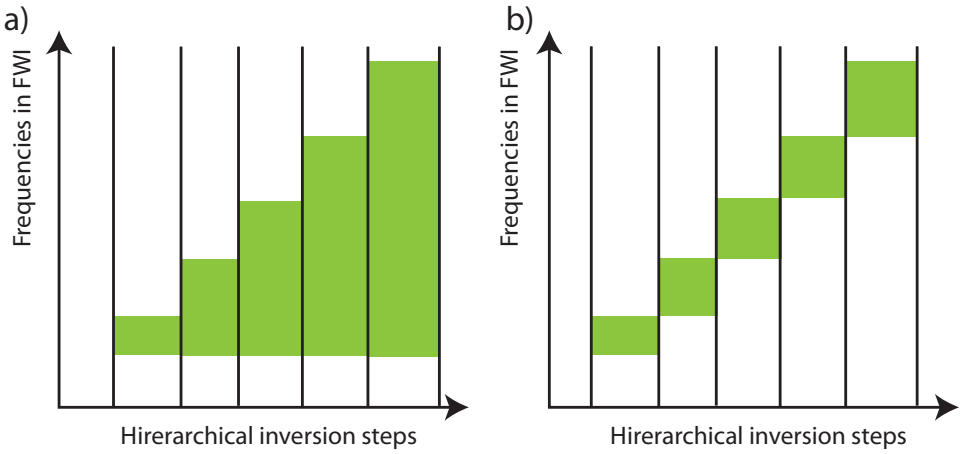

Figure 2.10: Two strategies to manage frequencies in FWI. (a) The frequency bandwidth is progressively broadened towards high frequencies during the different hierarchical steps. This approach is suitable for time-domain FWI, because only the highest frequency of the band governs the computational cost of the time-domain FWI (Bunks *et al.*, 1995). (b) Non-overlapping frequency groups of increasing high-frequency content are successively inverted (Brossier *et al.*, 2009). In (b), much less frequencies are inverted compared to (a), although the same overall frequency bandwidth is considered for FWI. This has important implications in term of computational cost if seismic modeling is performed in the frequency domain, the cost of which scales linearly with the number of frequencies.

where α_{min} is a constant, it depends on the offset and the reflector depth.

2.2.2 Hybrid FWI: time-domain modeling and frequency-domain inversion

We design a FWI algorithm in the frequency domain based on time domain modeling (Nihei et Li, 2007; Sirgue *et al.*, 2008). Our motivation behind the use of time domain modeling mainly relies on the scalability of time-domain modeling on large-scale computational platform, and the small memory demand of single-source modeling. These features provide some flexibility to implement time-domain modeling on computational platform of different architectures. A significant drawback is that attenuation cannot be implemented in the forward problem as easily as in the frequency domain. As we discussed in chapter 1.5, time-domain modeling can be parallelized with two levels of parallelism, which can be jointly used during one simulation by using two Message-Passing-Interface (MPI) communicators. The first level of parallelism distributes the source over the MPI processes, while the second level of parallelism relies on a domain decomposition of the computational mesh. The two levels of parallelism can be jointly used when the number of sources is much smaller than the number of processors, a possible configuration when source encoding techniques are used (Krebs *et al.*, 2009; Ben Hadj Ali *et al.*, 2011), or when a single-source simulation requires more memory than the one available on a shared-memory node. Furthermore, time windowing for selecting or muting some events in the seismograms (early arrivals, reflections, surface waves) is easy to implement in the time domain, unlike the frequency domain. We extract the frequency response by discrete Fourier transform (DFT) following the approaches promoted by Nihei et Li (2007) and Sirgue *et al.* (2008). One

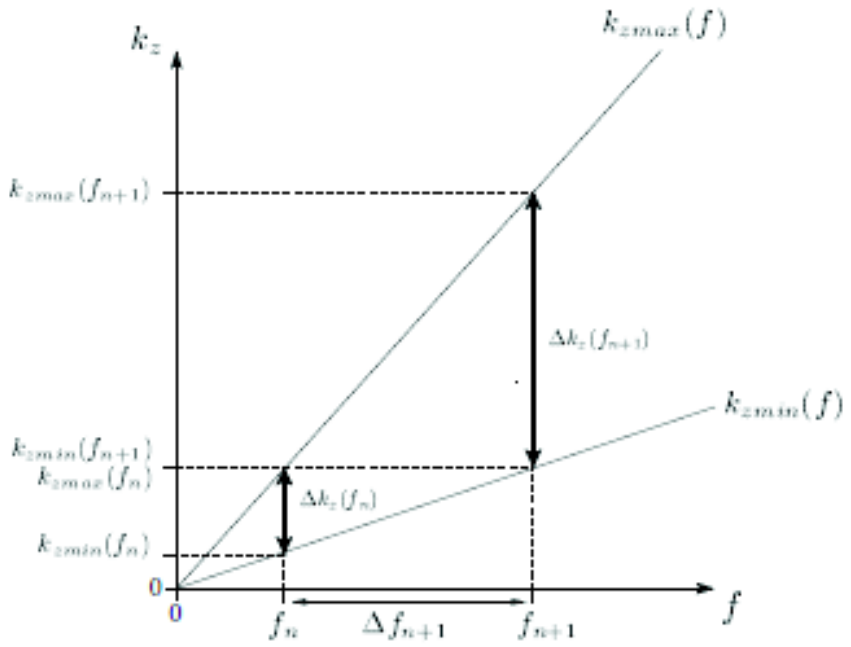

Figure 2.11: Illustration of the frequency discretization strategy (Sirgue et Pratt, 2004). $k_{zmin} = 4\pi f \alpha_{min}/c_0$ with $\alpha_{min} = 1/\sqrt{1+R^2}$ and $k_{zmax} = 4\pi f/c_0$, R is the maximum half offset-to-depth ratio, f is the frequency expressed in Hz and c_0 is the velocity in the background medium.

advantage of the approach based on the DFT is that an arbitrary number of frequencies can be extracted on the fly in the loop over time steps without significant extra computational cost. The phase sensitive detection approach of Nihei et Li (2007), where the source excitations are monochromatic, also allow one to compute the frequency response of multiple sources from one single time-domain modeling, by encoding sources with slightly different frequencies. Our FWI implementation is outlined in algorithm 2.1. As mentioned in the previous chapter, we compute the gradient of the misfit function with the adjoint state method from a pseudo-conservative form of the wave equation. This allows us to make the scattering kernel of the gradient independent of the differential operators, and hence of the numerical scheme used for seismic modeling. Therefore, different modeling engines on different meshes can be easily interfaced with the inversion. However, this requires on the one hand to project the wavefield solutions on the mesh that parameterizes the subsurface model during inversion, and on the other hand, to project the updated subsurface model on the mesh of the forward modeling (Figure 2.12).

Algorithm 2.1 The principle of the FWI algorithm

for i_freq_group = 1 to nb_frequency_group **do**
  for i_iter = 1 to nb_max_iteration **do**
    for i_src = 1 to nb_sources **do**
      ⇒ *computing the incident wavefield in time domain with the parallelism of the domain decomposition*
      for i_time = 1 to nb_time_step **do**
        computing the wavefield at time step i_time
        update the frequency solutions by DFT ∀ freq. ∈ i_freq_group
      end for
      ⇒ **projection the frequency solutions (forward problem towards inverse problem)**
      ⇒ *computing the residuals*
      ⇒ *computing the retro-propagate field in time domain with the parallelism of the domain decomposition*
      for i_time = nb_time_step to 1 **do**
        computing the wavefield at time step i_time
        update the frequency solutions by DFT ∀ freq. ∈ i_freq_group
      end for
      ⇒ **projection the frequency solutions (forward problem towards inverse problem)**
      ⇒ *update gradient with the expression* (2.29) ∀ freq. ∈ i_freq_group
    end for
    ⇒ *update the model with equation* (2.35)
    ⇒ **projection the parameters (inverse problem towards forward problem)**
  end for
end for

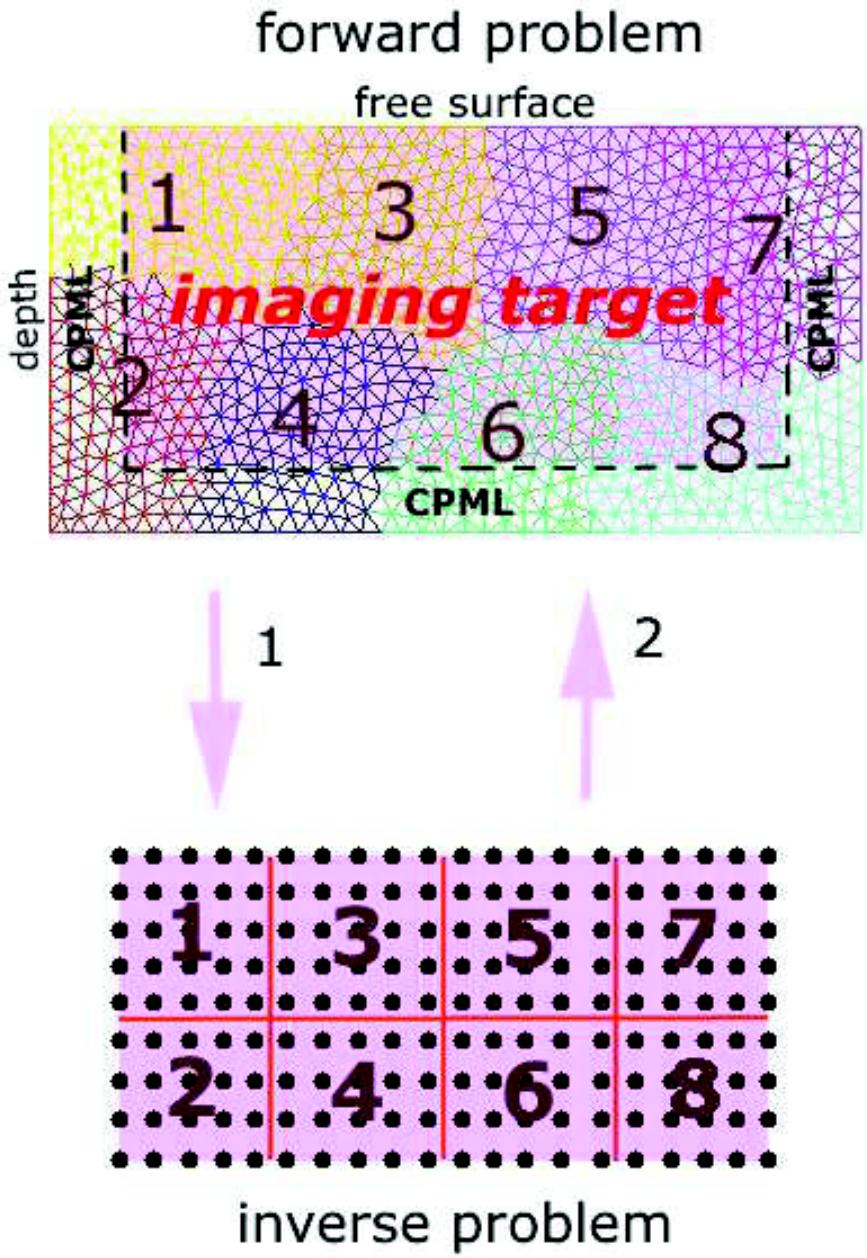

Figure 2.12: Illustration of the approach that consists in keeping independent the parameterizations used for the forward and inverse problems. The triangular computational mesh is subdivided into eight subdomains for seismic modeling. This mesh includes the CPML layers, which are not considered in the inverse problem. The Cartesian grid that is used to solve the inverse problem is also subdivided in eight subdomains (ideally, we use the same number of processors to perform the forward and inverse problems). The wavefield solutions of the forward problem are projected onto the inversion grid to solve the inverse problem, while the subsurface model updated by FWI is projected on the modeling mesh before moving to the next iteration Etienne *et al.* (2010).

2.2.3 Parallelism over shots versus domain decomposition

The main objective of seismic full waveform inversion for the petroleum industry is to obtain a high resolution velocity model, which will be used as a macro-model for migration and, subsequently, for seismic interpretation. Three-dimensional FWI is feasible today thanks to the increase of the available computational power and some algorithmic enhancements, but its application still remains limited at low frequencies (lower than 10 Hz (Sirgue *et al.*, 2009)), because of the prohibitive computational cost. The main computational burden results from multi-source modeling. Usually, thousands of sources are used in three-dimensional surveys (i.e., 2302 sources in the Valhall case study; in fact, this number represents the number of receivers, which are processed as sources in virtue of the reciprocity of Green functions). This complexity can be reduced with different strategies. We present two of these strategies.

One strategy is to speed up the forward problem by using a parallel programming structure. The domain decomposition of the physical domain will give a first level of parallelism. We can use distributed memory architecture based on MPI primitives and, on fat nodes, shared memory architecture could improve the speedup of the forward modeling.

Another strategy is modeling different shots using an embarrassing parallel strategy. The second level of parallelism does not require any communication between the different shots. Only at the end of the computation of the incident and adjoint wavefields, the gradients associated with distinct sources are stacked through collective communications. Figure 2.13 shows the speedup obtained by using the parallelism over shots. We perform modeling for 64 sources in a homogeneous infinite medium ($101 \times 201 \times 201$). The test was performed on the cluster Licallo hosted by Observatoire de la Cote d'Azur, and shows a quite good speedup as expected.

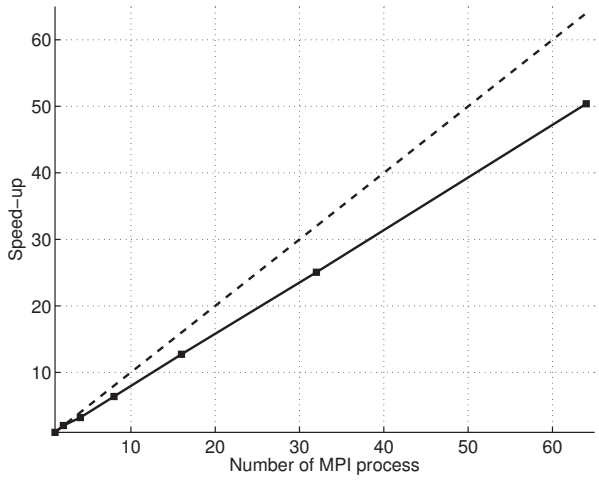

Figure 2.13: The speedup for parallelism over the shots.

Combining the two level parallelisms provides an additional level of flexibility to optimize the performances of the codes as a function of the problem size and computational architecture. Figure 2.14 illustrates how the two levels of parallelism can be combined, through the use of three MPI communicators (the global one plus two other ones, which are dedicated to specific tasks described hereafter).

Table 2.1: Comparison between the different parallelisms.

Parallelism	subdomain	Number of parallel sources	memory	time
Domain decomposition	64	1	16.06 MB	567.6525 s
Parallelism over sources	1	64	436.56 MB	287.8543 s
Double level parallelism	2	32	217.96 MB	283.8729 s

The first level of parallelism is used to perform single-source modeling in parallel on a group of processors using domain decomposition of the computational domain. Point-to-point communications are required to exchange wavefield values between close subdomains, which share an overlapping band of grid nodes. The sources of the experiment are distributed on groups of processors, and the single-source modelings are independently performed on each group of processors. Once the incident and adjoint wavefields are computed on a group of processor, the gradient of the misfit function associated with this source is computed in parallel in distributed form over the sub-domains. The second level of parallelism aims to stack the gradients associated with each source through collective communications on each subdomain of the computational mesh. This strategy is illustrated in figure 2.14 where two single-source modelings are performed on two groups of fours processors. The table 2.1 shows the time and memory required for a subsurface model of dimension $101 \times 201 \times 201$) using 64 processors. Note that the parallelism over the shots shows a more efficient speedup than domain decomposition, but the required memory per MPI process is more expensive. Parallelism based on a domain decomposition of the physical domain can efficiently reduce the memory complexity (Virieux et Operto, 2009). The combination the two level parallelism provides the best trade-off between computing time and memory demand.

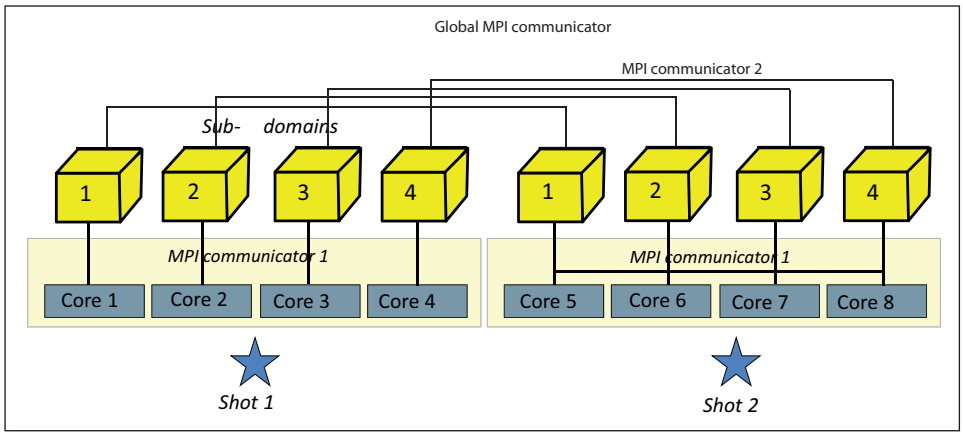

Figure 2.14: Illustration of the two-level parallelisms for 2 shots and, the domain decomposition performed on 4 processors.

2.3 Validation of the FWI code with synthetic examples

2.3.1 The synthetic channel model

In this section, I present the numerical experiments that we design to validate our full waveform inversion (FWI) code. These numerical tests use a small target of the EAGE/SEG Overthrust model (Aminzadeh *et al.*, 1997) centered on a channel to keep the computational cost tractable on the computational platforms that were available to us. Three tests are presented in this section. The first test corresponds to an onshore model with an absorbing boundary condition on top of the model. Therefore, no free-surface multiples are involved in the inversion. In the second test, the same onshore model is used but a free surface is set on top of the model. In the third test, we add a shallow water layer and a soft sedimentary layer on top of the model to mimic a marine experiment similar to the real data case study presented in the final section of this report. A free surface is used on top of the model to involve free-surface multiples in the inversion, a critical issue in marine environment.

The channel target covers an area of 2.25 $km \times 7.4\ km \times 9.0\ km$ in the directions of depth, cross-line, and dip-line, respectively[1]. The model is discretized with a grid spacing of 50 m (Figure 2.15). It leads to a $46 \times 149 \times 181$ grid. The minimum and maximum velocities are 3389 m/s and 5500 m/s, respectively. 21×20 sources are located nearby the surface (source depth depends on the different tests) with spacing of 350 m and 450 m in the cross-line and dip-line directions, respectively. 75×91 receivers are uniformly deployed at the same depth than sources with a spacing of 100 m. The starting velocity model is obtained by smoothing the true model with a three-dimensional (3D) Gaussian function (the correlation length is 250 m in the three directions). The channel structure and the vertical layers are almost non visible in the starting model (Figure 2.16) and the goal of the inversion is to reconstruct them with the highest possible resolution. The maximum frequency involved in the inversion is 12 Hz. The frequency groups are defined depending on the different tests. For each frequency group, we computed 10 iterations. Let us focus on the FWI results in the following sections.

2.3.2 Onshore model without free surface

The first test considers an onshore model. No free surface multiples are involved in the inversion: an absorbing boundary condition is implemented on top of the true model and of the FWI models. The PMLs contain 5 nodes in all of the directions. Both sources and receivers are located in a horizontal plane buried at a depth of 150 m. FWI is applied successively to two frequency groups composed of 5 frequencies, [3.29 3.76 4.15 4.59 5.025] Hz, for the first group, and 15 frequencies, [5.025 5.23 5.454 5.66 5.87 6.09 6.53 6.97 7.41 7.85 8.29 8.73 9.61 10.27 11.14 12.01] Hz, for the second group. The observed data is computed in the true model with a grid spacing of 50 m. The FWI set-up and the computational time are shown in the table 2.2 for the two frequency groups. The parallelism is performed by shot distribution over processors. The density is assumed to be constant and is set to 1000 $kg.m^{-3}$. We assume that the source wavelet is not known, and therefore it is estimated during FWI. The inversion is performed on 420 cores of the cluster Babel[2]. Figure 2.17 shows the FWI results for the first frequency group. The FWI succeeds in reconstructing the different layers of the model as well as the channel with a resolution that is consistent with the inverted frequencies. A square pattern is

[1] It is extracted from the EAGE/SEG Overthrust model: from 1175 m to 3425 m in the vertical, form 9225

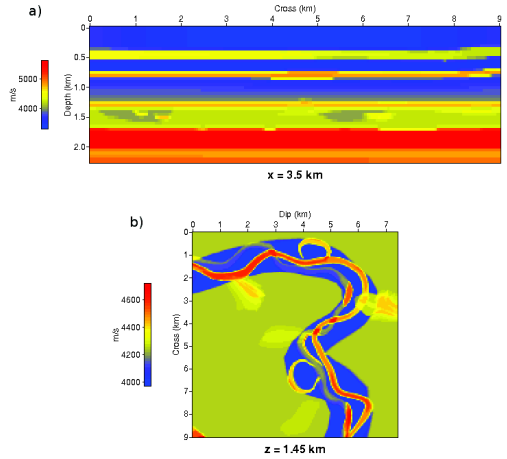

Figure 2.15: Overthrust model - True model. (a) vertical section at x = 3.5 km; (b) horizontal section at z = 1.45 km.

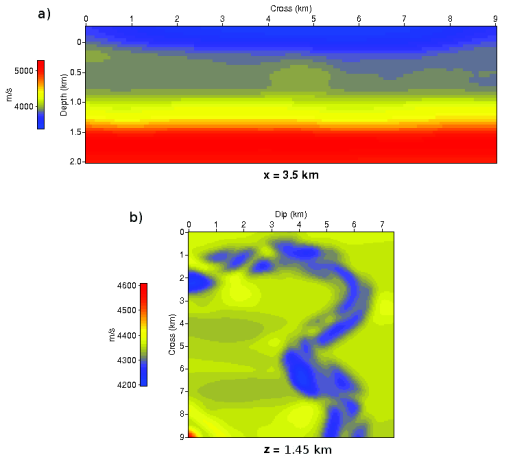

Figure 2.16: Overthrust model - initial model for FWI. (a) vertical section at x = 3.5 km; (b) horizontal section at z = 1.45 km.

Table 2.2: Onshore model without free surface - FWI set-up: N_{gre} denotes the inverted frequency group; The dimensions of the model are denoted by N_z, N_x, and N_y in the three directions; Δt denotes the time step of the simulation; The grid interval Δh is identical in the three directions; F (Hz) denoted the frequency band; The number of source and receiver are denoted by N_{sour} and N_{rec}, respectively; N_{core} denotes the number of processor; N_{ite} is the number of iteration; The computational time for the inversion is denoted by T (h); The computation is performed on the cluster Babel[2]

N_{gre}	N_z	N_x	N_y	Δt	Δh (m)	$F(Hz)$	N_{sour}	N_{rec}	N_{core}	N_{ite}	T (h)
First group	46	149	181	0.0035	50	3.29–5.03	420	6825	420	10	15.4
Second group	46	149	181	0.0035	50	5.23–12.01	420	6825	420	10	19.6

superimposed on the channel (Figure 2.17c) in the FWI model, and represents the footprint of the acquisition. This footprint will be reduced by destructive interferences as the inversion progresses towards high frequencies. The misfit function versus the iterations is shown in figure 2.18. The convergence was not reached after 10 iterations, and a more complete reconstruction of the amplitude of the model perturbations would have been obtained with more iterations. We use the final model of the first frequency group inversion (Figure 2.17(b,e)) as the starting model for the second frequency group. The FWI set-up and computational time are shown in table 2.2. Figure 2.19 shows the comparison between the true model and the final FWI model. Note that our maximum frequency involved in this group goes up to 12 Hz, but the resolution

m to 16625 m in the cross direction and from 5725 m to 14725 m in the dip direction.

[2]IBM Blue Gene/P system. It has 10 racks, each one containing 1024 compute nodes. A compute nodes has four computing cores running at 850MHz and 2GiB of memory. The total theoretical peak performance is 139Tflops (3.4Gflops by core)

of the FWI model is not as high as the one corresponding to this frequency: this vertical resolution should be around 140 m (Figure 2.19g). This results because we simultaneously invert frequencies between 5 Hz to 12 Hz without any data preconditioning. As the source is a Ricker of central frequency 5 Hz, the 5 Hz frequency has a much higher weight in the misfit function than the 12 Hz frequency (the amplitude spectrum of the source wavelet acts as a data preconditioning in the misfit function, see figure 2.22). Let's remind that the gradient of the misfit function can be viewed as an inverse Fourier transform, where the coefficients of the Fourier series are weighted by the different mono-frequency data residuals. Therefore, the amplitude variations with frequency of the mono-frequency data residuals act as a bandpass filtering on the gradient of the misfit function, when multiple frequencies are jointly inverted.

Note also how the footprint of the acquisition in figure 2.19c was significantly reduced compared to the one shown in figure 2.17c. For completeness, I show a horizontal plane of the FWI model, which is located above the plane of sources and receivers (Figure 2.20). In this plane, the footprint of the acquisition geometry is not visible.

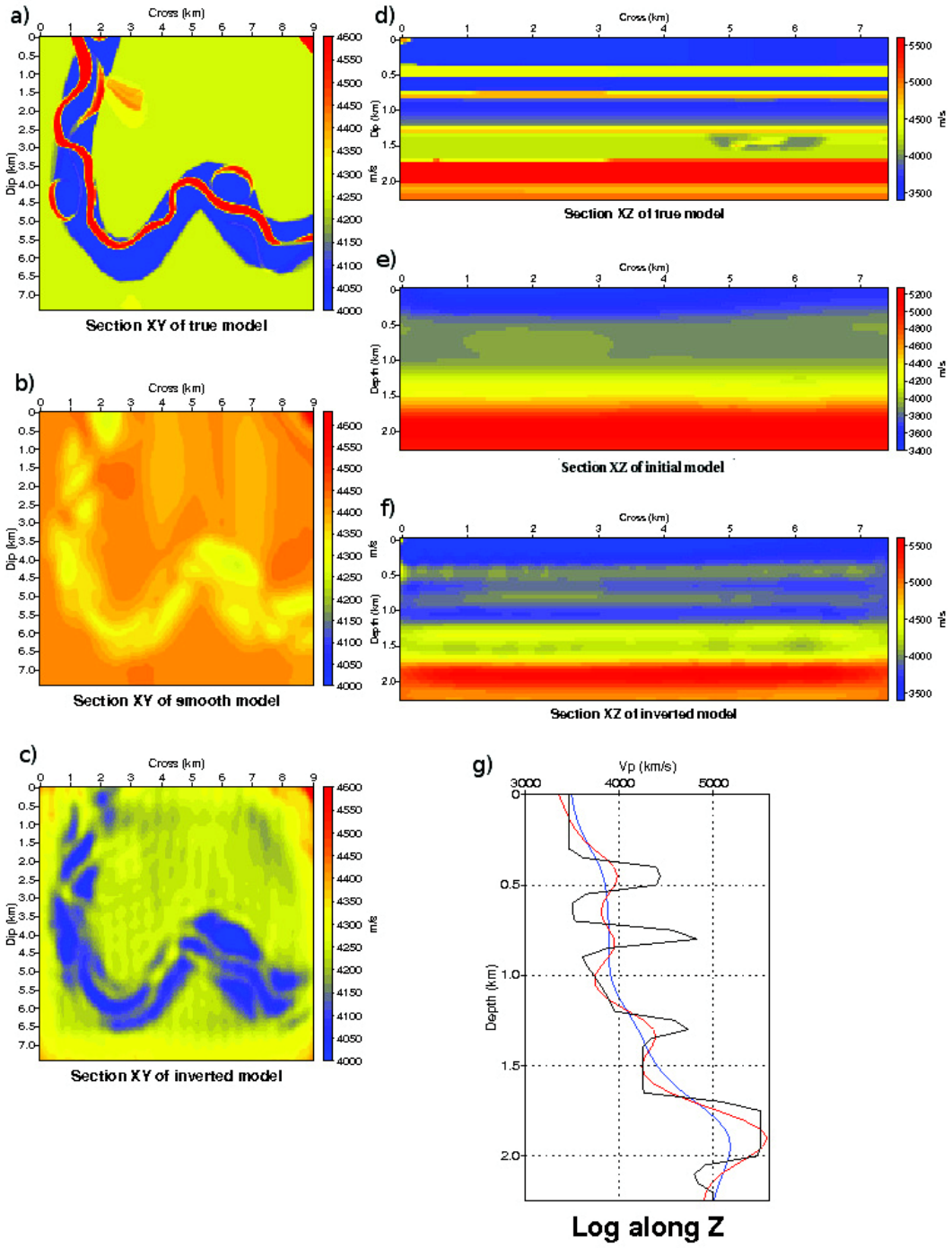

Figure 2.17: Onshore model without free surface - FWI results for the first frequency group inversion. (a, d) vertical (y = 3.5 km) and horizontal (z = 1.45 km) sections of the true model, respectively. (b, e) Same as (a, d) for the initial model. (c, f) Same as (a, d) for the FWI model. (g) Vertical profile extracted in the middle of the model (x= 4.5 km, y= 3.5 km). The blue, black, and red curves are from the initial model, the true model, and the FWI model, respectively.

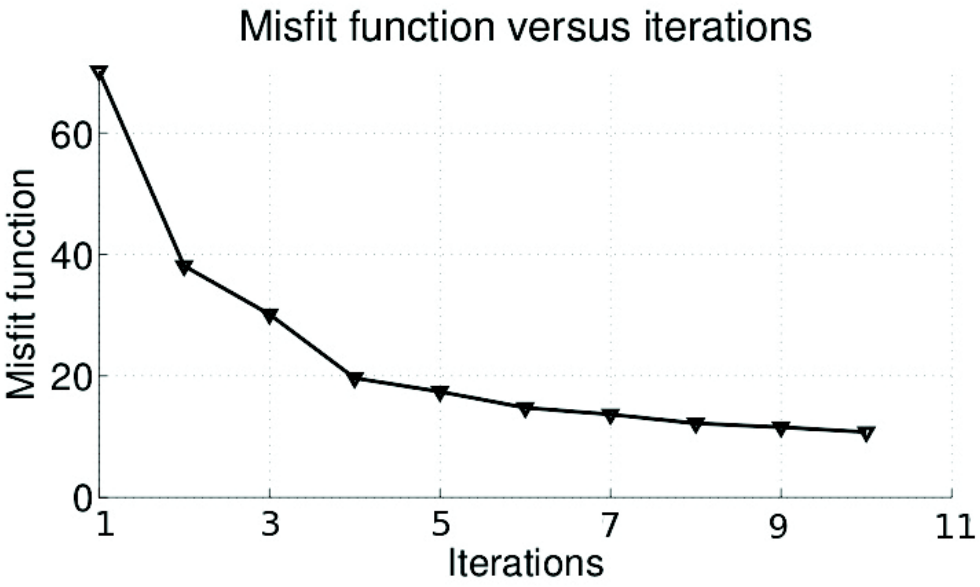

Figure 2.18: Onshore model without free surface - Misfit function versus iterations for the first frequency group inversion. The number of iterations is ten. Note that more iterations could have been performed to more accurately recover the amplitude of the model perturbations.

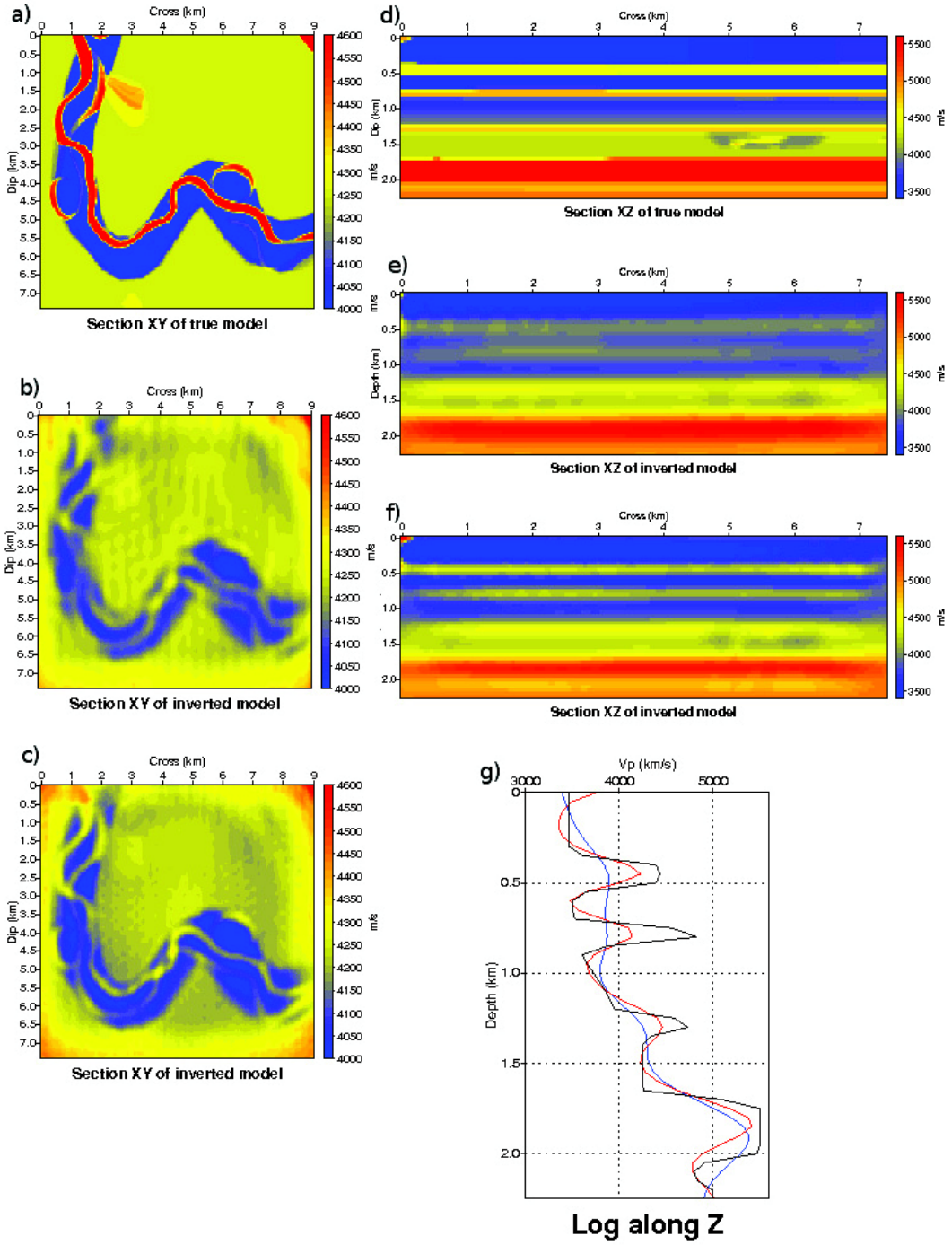

Figure 2.19: Onshore model without free surface - FWI results for the second frequency group inversion. Same as figure 2.17 for the frist frequency group inversion. Note how the resolution was improved compared to the FWI model of the first frequency group inversion (Figure 2.17c). Note also how the acquisition footprint was reduced by stacking the contribution of higher frequencies.

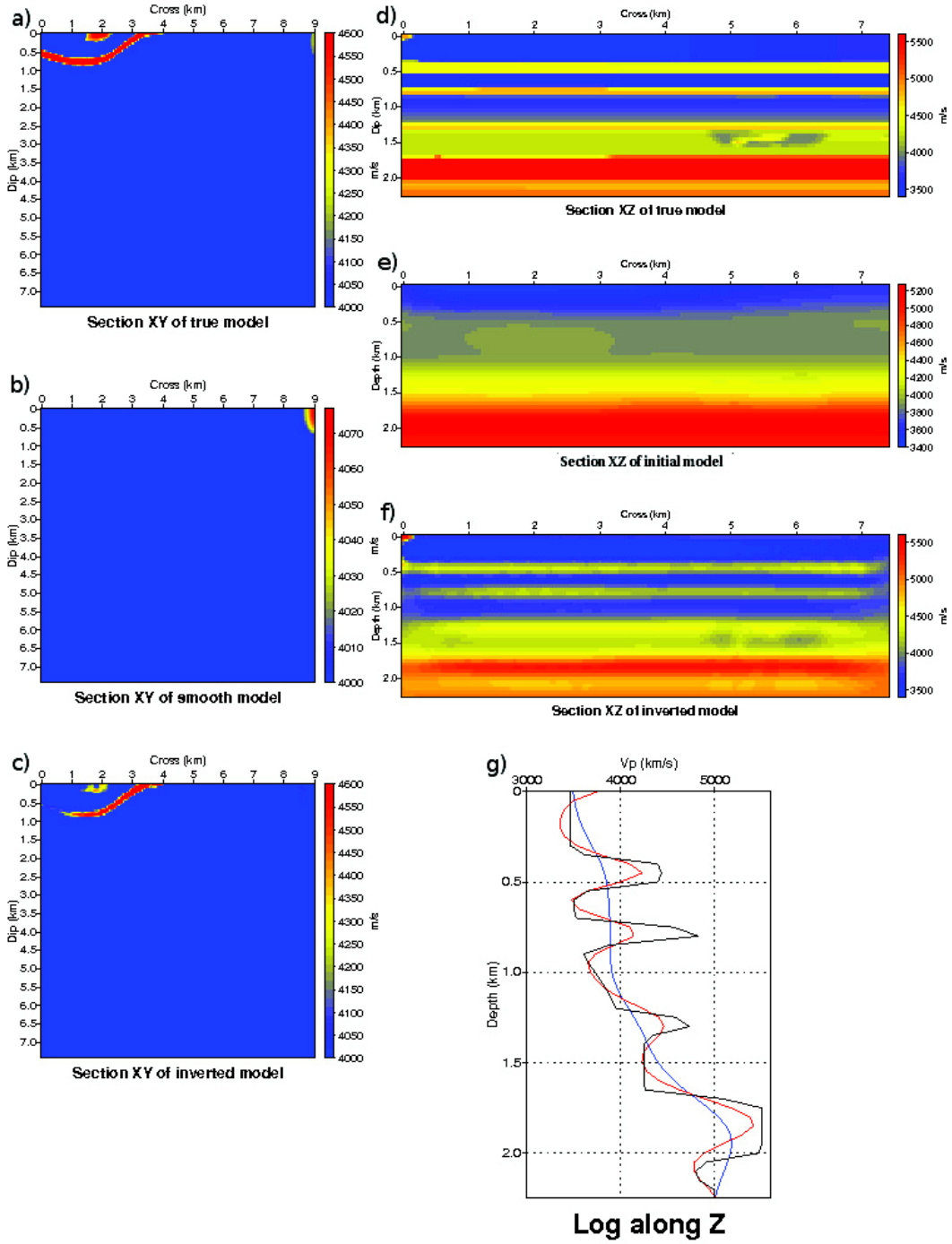

Figure 2.20: Onshore model without free surface - FWI results for the second frequency group inversion. Same as figure 2.19 for the second frequency group inversion but the horizontal section of the FWI model is extracted above the source-receiver plane at the depth of 100 m. Note the acquisition footprint is not visible.

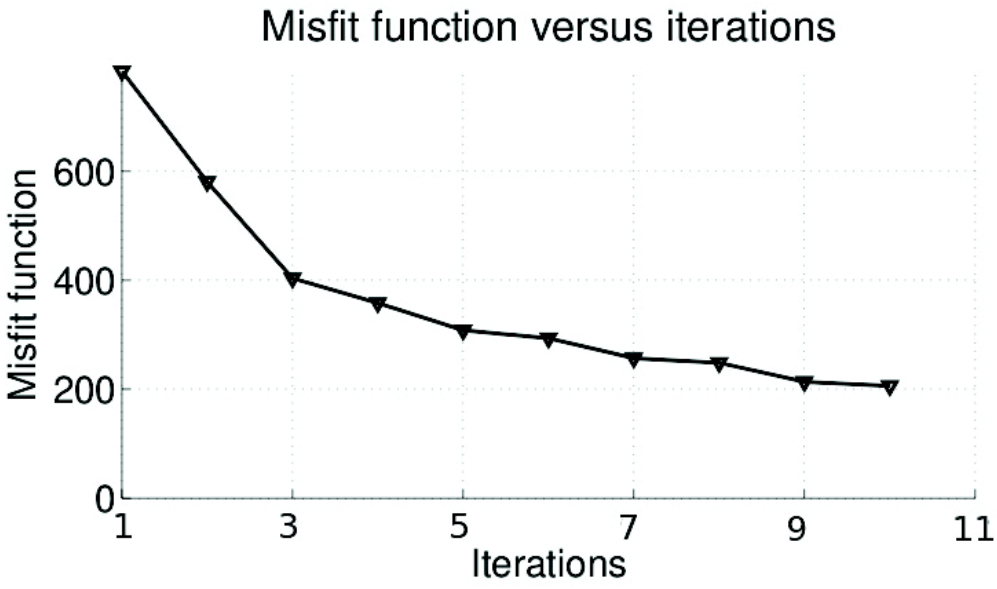

Figure 2.21: Onshore model without free surface - Misfit function versus iterations for the second frequency group inversion. The maximum number of iterations is 10.

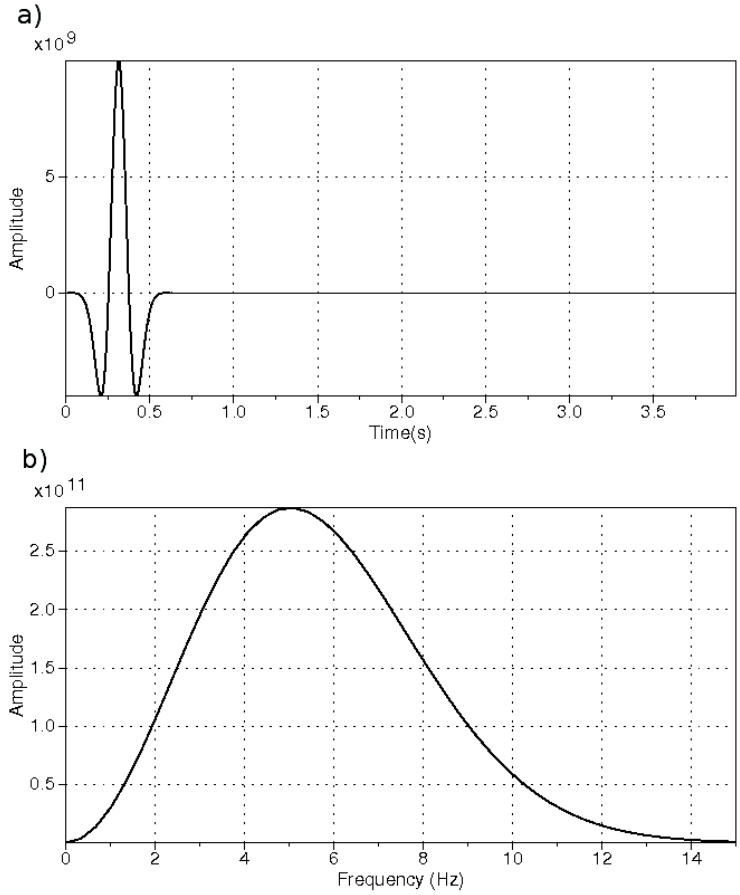

Figure 2.22: Source wavelet - Ricker wavelet of central frequency 5 Hz (a) and its spectrum (b). Note the small spectral amplitude of the 12 Hz frequency compared to that of the 5 Hz frequency. The amplitude spectrum acts as a data weighting in the misfit function, which translates into a bandpass filtering of the gradient of the misfit function.

2.3.3 Onshore model with free surface

To mimic a more realistic survey, we set a free surface on top of the previous onshore model. The sources and receivers are located at 70 m and 5 m in depth, respectively. We use three frequency groups composed of five frequencies [3.29 3.76 4.15 4.59 5.025] Hz for the first group, and eight frequencies [5.025 5.23 5.454 5.66 5.87 6.09 6.53 6.97] Hz for the second frequency group, and nine frequencies [6.97 7.41 7.85 8.29 8.73 9.61 10.27 11.14 12.01] Hz for the third group. We use three frequency groups instead of two in the previous test, because free surface multiples are expected to increase the non linearity of the FWI. The FWI set-up and computational time are shown in table 2.3. The grid interval is 50 m for each frequency groups to prevent numerical dispersion for the maximum frequency. As the sources and receivers does not coincide with the grid nodes, the *Hicks* interpolation is used. Ten iterations are performed per frequency group for a total of 30 iterations.

Table 2.3: Onshore model with free surface - FWI set-up: N_{gre} denotes the inverted frequency group; The dimensions of the model are denoted by N_z, N_x, and N_y in the three directions; Δt denotes the time step of the simulation; The grid interval Δh is identical in the three directions; F (Hz) denoted the frequency band; The number of source and receiver are denoted by N_{sour} and N_{rec}, respectively; N_{core} denotes the number of processor; N_{ite} is the number of iteration; The computational time for the inversion is denoted by T (h); The computation is performed on the cluster Babel

N_{gre}	N_z	N_x	N_y	Δt	Δh (m)	$F(Hz)$	N_{sour}	N_{rec}	N_{core}	N_{ite}	T (h)
First group	46	149	181	0.0035	50	3.29–5.03	420	6825	420	10	12.7
Second group	46	149	181	0.0035	50	5.02–6.97	420	6825	420	10	14.4
Third group	46	149	181	0.0035	50	6.97–12.01	420	6825	420	10	19.9

The figure 2.23 shows the FWI results of the first frequency group with the same kind of representation than for the previous test. Note that the footprint acquisition can be seen more clearly as well in the form of stripes in the channel (Figure 2.23c). Nonetheless the shape of the channel and the vertical layers are well imaged in this case with free surface. The misfit function versus the iterations is shown in figure 2.24. It is worth noting that the vertical resolution of the FWI model of this test seems higher than that of the model of the previous test (compare figures 2.17 and 2.23). This might results from the higher illumination provided by the free-surface multiples in the data. The misfit function versus the iterations is shown in figure 2.24.

We successively applied our FWI algorithm to the second and third frequency groups. The computing statistics are shown in table 2.3. The FWI results are shown in figures 2.25 and 2.27 for second and third frequency groups, respectively. Figure 2.28c shows the imaging nearby the free surface at a depth of 100 m. Compared to the case of the model without free surface (Figure 2.20c), we obtain a similar result. However, the footprint of the acquisition is more significant when there is a free surface on top of the model. Results of the third frequency group inversion confirms that the final FWI model (Figure 2.27c) has a better resolution in depth than the final FWI model inferred from the data set without free-surface multiples (Figure 2.19c). Again, we believe that this improvement results, because free-surface multiples can help to increase the subsurface illumination. Another possible reason is that we narrow the band of the last frequency group (7 Hz to 12 Hz for this test instead of 5 Hz to 12 Hz for the previous test). The 12 Hz frequency has a higher weight during inversion of the [7 Hz - 12 Hz] frequency group than during inversion of the [5 Hz - 12 Hz] one, because of the spectral amplitudes of the 5 Hz Ricker wavelet. Figure 2.26 and 2.29 show the misfit function versus the iterations.

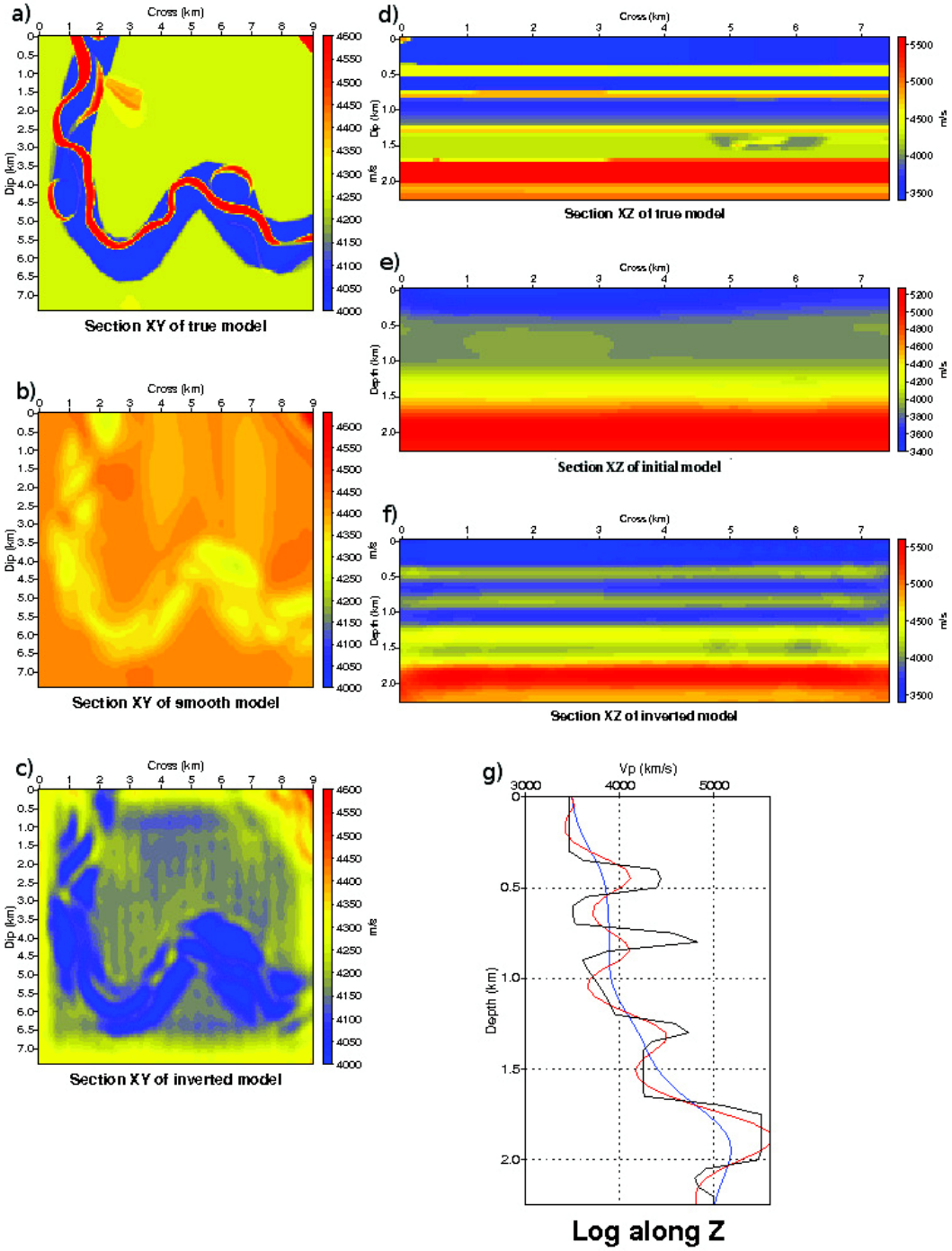

Figure 2.23: Onshore model with free surface - FWI results for the first frequency group inversion. (a, d) vertical (y = 3.5 km) and horizontal (z = 1.45 km) sections of the true model, respectively. (b, e) Same as (a, d) for the initial model. (c, f) Same as (a, d) for the FWI model. (g) Vertical profile extracted in the middle of the model (x= 4.5 km, y= 3.5 km). The blue, black, and red curves are from the initial model, the true model, and the FWI model, respectively.

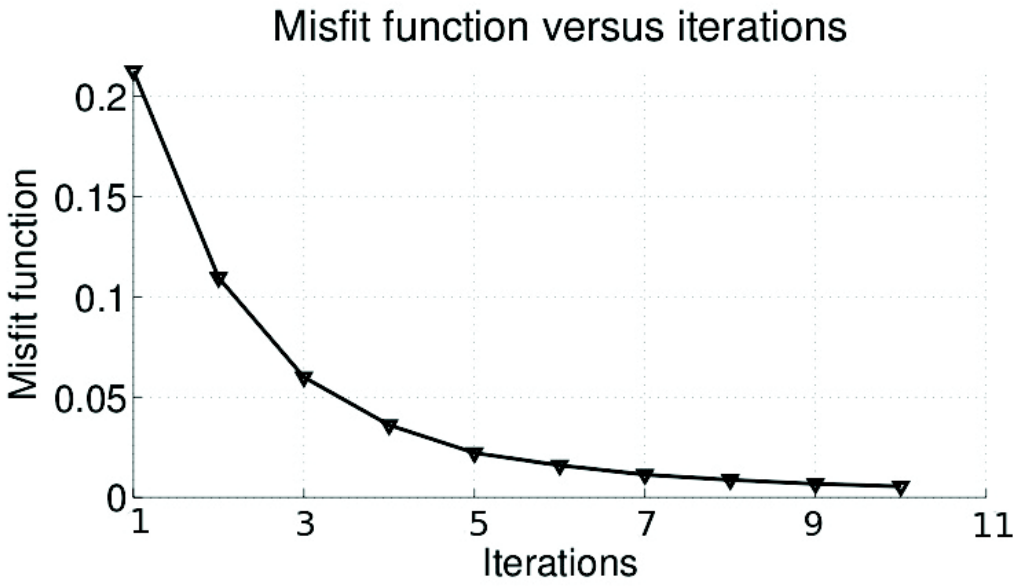

Figure 2.24: Onshore model with free surface - Misfit function versus iterations for the first frequency group inversion. As for the previous test, more iterations could have been performed to improve the reconstruction of the amplitudes of the model perturbations.

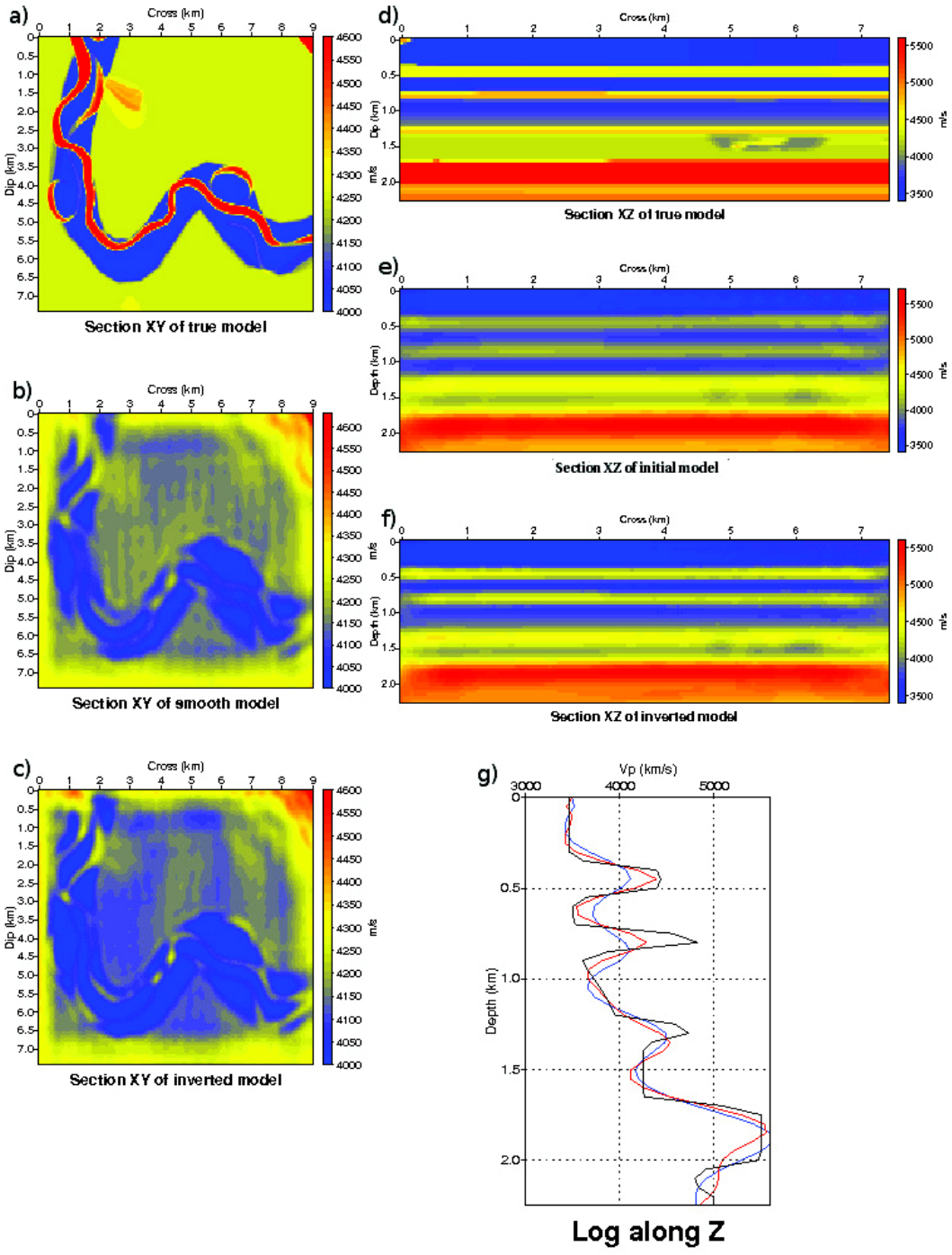

Figure 2.25: Onshore model with free surface - FWI results for the second frequency group inversion. Same as figure 2.23 for the second frequency group inversion.

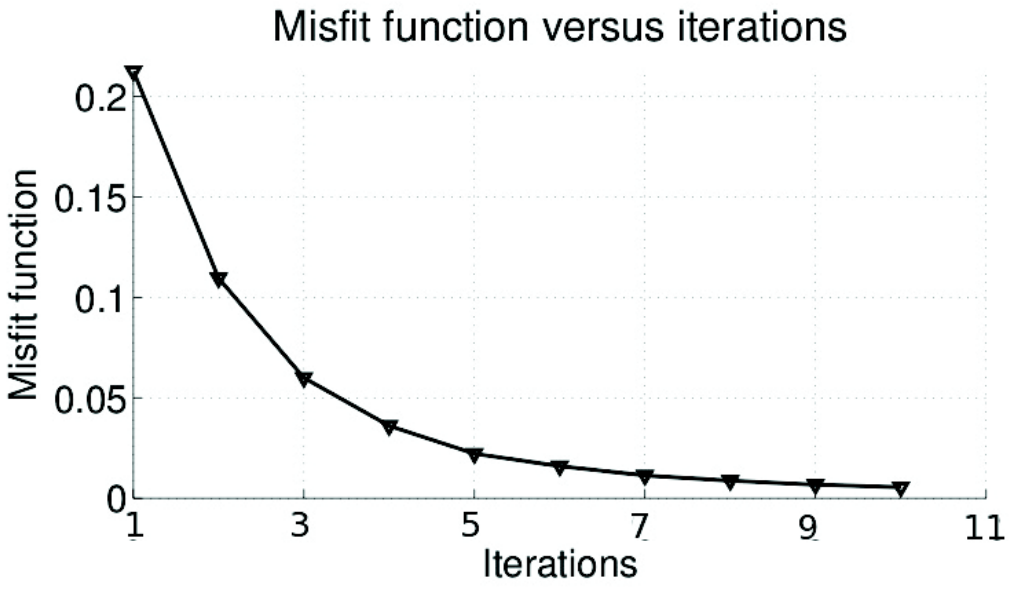

Figure 2.26: Onshore model with free surface - Misfit function versus iterations for the second frequency group inversion.

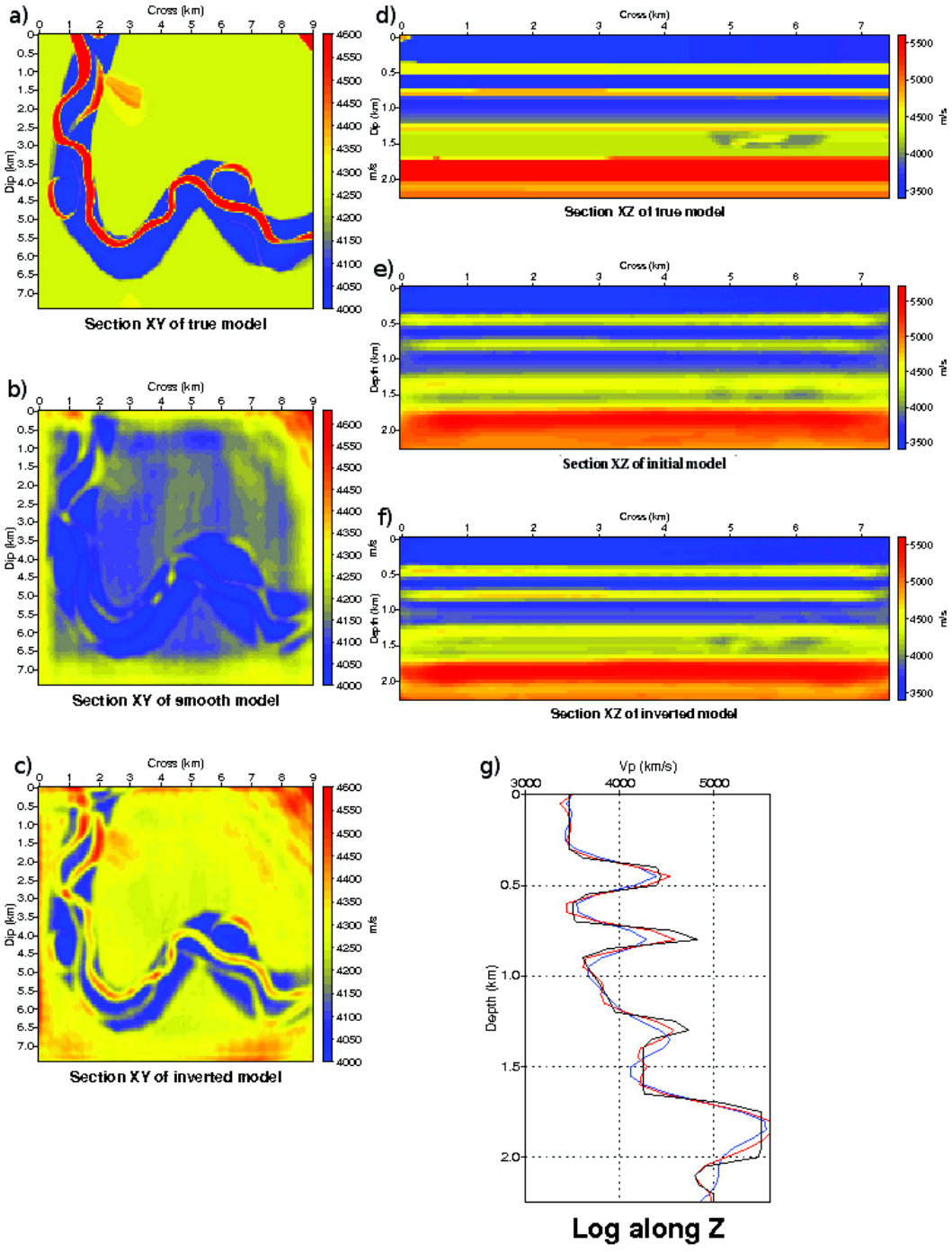

Figure 2.27: Onshore model with free surface - FWI results for the third frequency group inversion. Same as figure 2.23 for the third frequency group inversion.

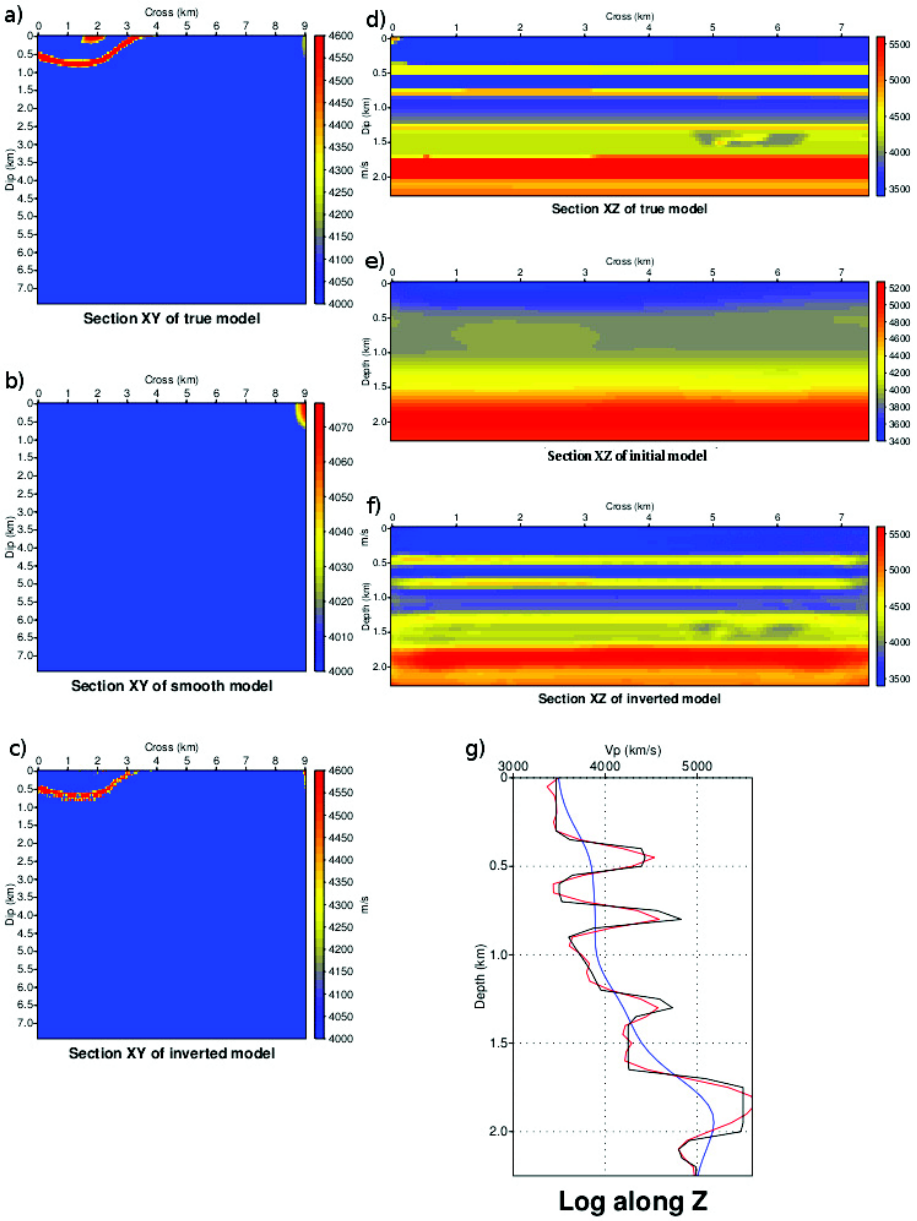

Figure 2.28: Onshore model with free surface - FWI results for the third frequency group inversion. Same as figure 2.27 but the horizontal section of the FWI model is extracted above the source-receiver plane at the depth of 100 m. Note the acquisition footprint is not visible.

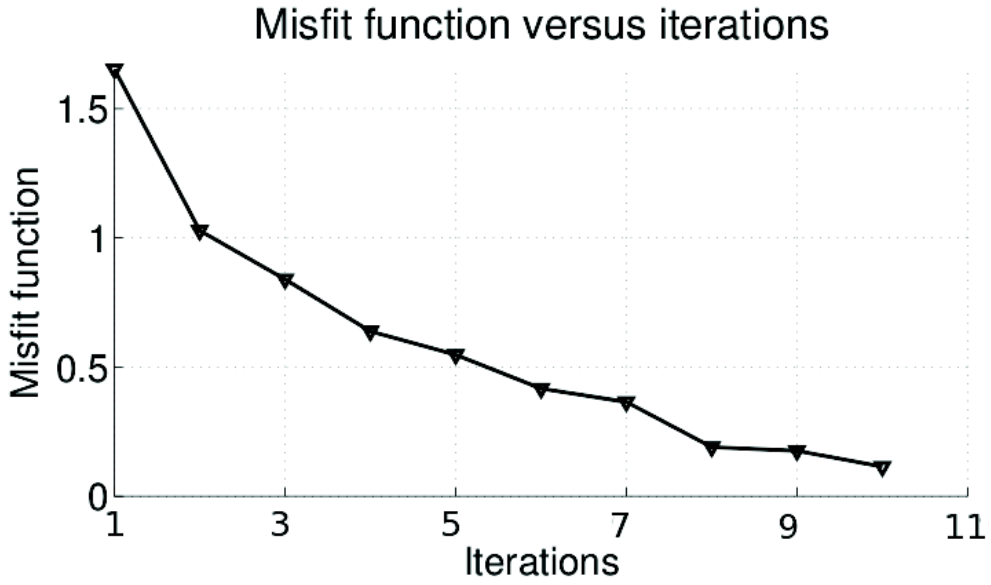

Figure 2.29: Onshore model with free surface - Misfit function versus iterations for the third frequency group inversion. More iterations could have been performed to reach the convergence, which is slower than for the case study without free surface. More complex wave phenomena, which result from the presence of free surface multiples, likely slow down the convergence of the FWI.

Table 2.4: Offshore model with free surface - FWI set-up and computational time N_{gre} denotes the inverted frequency group; The dimensions of the model are denoted by N_z, N_x, and N_y in the three directions; Δt denotes the time step of the simulation; The grid interval Δh is identical in the three directions; F (Hz) denoted the frequency band; The number of source and receiver are denoted by N_{sour} and N_{rec}, respectively; N_{core} denotes the number of processor; N_{ite} is the number of iteration; The computational time for the inversion is denoted by T (h); The computation is performed on the cluster Licallo.

N_{gre}	N_z	N_x	N_y	Δt	Δh (m)	$F(Hz)$	N_{sour}	N_{rec}	N_{core}	N_{ite}	T (h)
First group	34	100	121	0.009	75	3.29–5.02	420	6825	420	10	2.1
Second group	34	100	121	0.009	75	5.02–6.97	420	6825	420	10	2.6
Third group	100	298	361	0.003	25	6.97–12.01	420	6825	420	10	71.5

2.3.4 Offshore model with free surface

Application of FWI to marine data from the Valhall field is the final goal of this thesis. Therefore, we design a marine validation test using the EAGE/SEG Overthrust channel target model (Figure 2.15). We add a water layer and a soft sedimentary layer on top of the model (velocity is 1500 m/s). The sea bottom is at 75 m in depth. The thickness of the sedimentary layer is 150 m with velocities between 1700 and 1900 m/s. To avoid too high velocity contrasts between the soft sedimentary layer and the underlying layers, we scale the velocities of the channel target model with a factor of 0.65 (Figure 2.30). After this scaling, the velocities range between 1500 m/s and 3390 m/s in the channel target of the overthrust model. The initial model is obtained by smoothing the channel target model with a 3D Gaussian function of correlation length 250 m in the three directions (Figure 2.31). The observed data are computed in the true model with a grid spacing 25 m. Different grid spacings are used for the inversion: 75 m for the first two frequency groups and 25 m for the last group. We use the same strategy as for the onshore model with free surface. Three frequency groups are successively inverted: [3.29 3.76 4.15 4.59 5.025] Hz, [5.025 5.23 5.454 5.66 5.87 6.09 6.53 6.97] Hz, and [6.97 7.41 7.85 8.29 8.73 9.61 10.27 11.14 12.01] Hz. PMLs contain four nodes, and a free surface is on top of the model. A surface acquisition similar to the one of the Valhall experiment is designed: sources are located at 70 m in depth on the sea bottom, and receivers are just below the free surface at 5 m in depth. The source wavelet is estimated during FWI. This test is performed in Licallo[3]. Note the computation is $5 \sim 6$ times faster on Licallo than on Babel (which is we used for the previous tests.) Ten iterations are performed per frequency group. Results of the first frequency group inversion are outlined in figure 2.32. The FWI computing statistics are outlined in table 2.4. The channel target is not as well imaged as for the onshore model. This might result from the shallow water layer, where a significant energy partioning can occur. Moreover, the acquisition footprint is much stronger in the marine case, compared to the onshore case study with free surface. The misfit function versus the iterations is shown in figure 2.33. We have successfully applied our FWI algorithm to the second and third frequency groups. The FWI computing statistics are shown in table 2.4. The FWI results are shown in figures 2.34 and 2.36 for the

[3]This mid-sized computing center hosted at O.C.A is the result of an joint effort initiated in 2004. This machine handles all administrative tasks in the cluster. It includes: Processor: 2 x Intel (R) Xeon (R) CPU E5620@2.40GHz (4 cores, hyper threading enabled) and Memory: 24GB

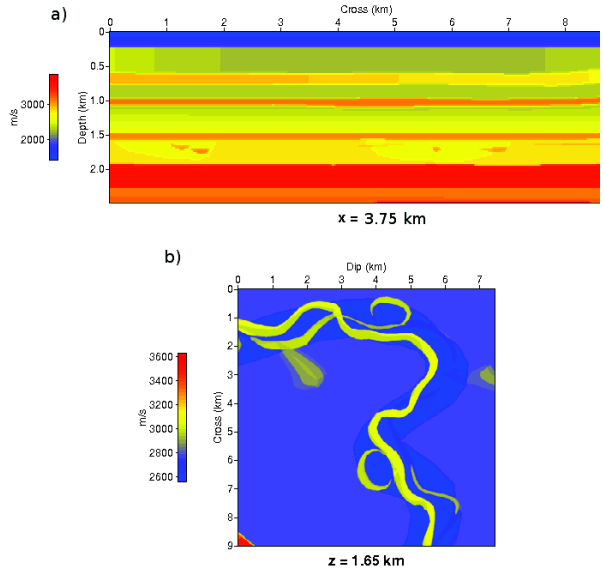

Figure 2.30: Offshore model with free surface - True model: a) a vertical section at x = 3.75 km. b) a horizontal section at z = 1.65 km.

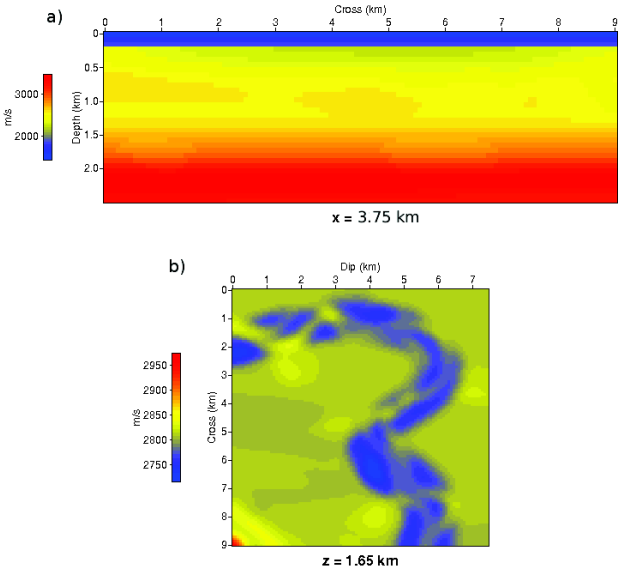

Figure 2.31: Offshore model with free surface - Initial model: a) a vertical section at x = 3.75 km. b) a horizontal section at z = 1.65 km.

second and third groups, respectively.

The final FWI model (figure 2.36c and f) shows a more accurate reconstruction of the velocity amplitudes than for the onshore model. This might result, because the velocity contrasts were decreased in the offshore model relatively to the onshore one. A more detailed evaluation of the result accuracy is shown in the velocity profiles shown in figure 2.36g. We note that, in the near surface, the shape of the geological structures are not as well imaged as in depth, probably because of the footprint of the acquisition (under 1.5 km in figure 2.36g). The misfit function versus the iterations are shown in figures 2.35 and 2.37 for the second and third groups, respectively. In both cases, a good convergence is achieved, although more iterations would have allowed to improve the velocity reconstruction.

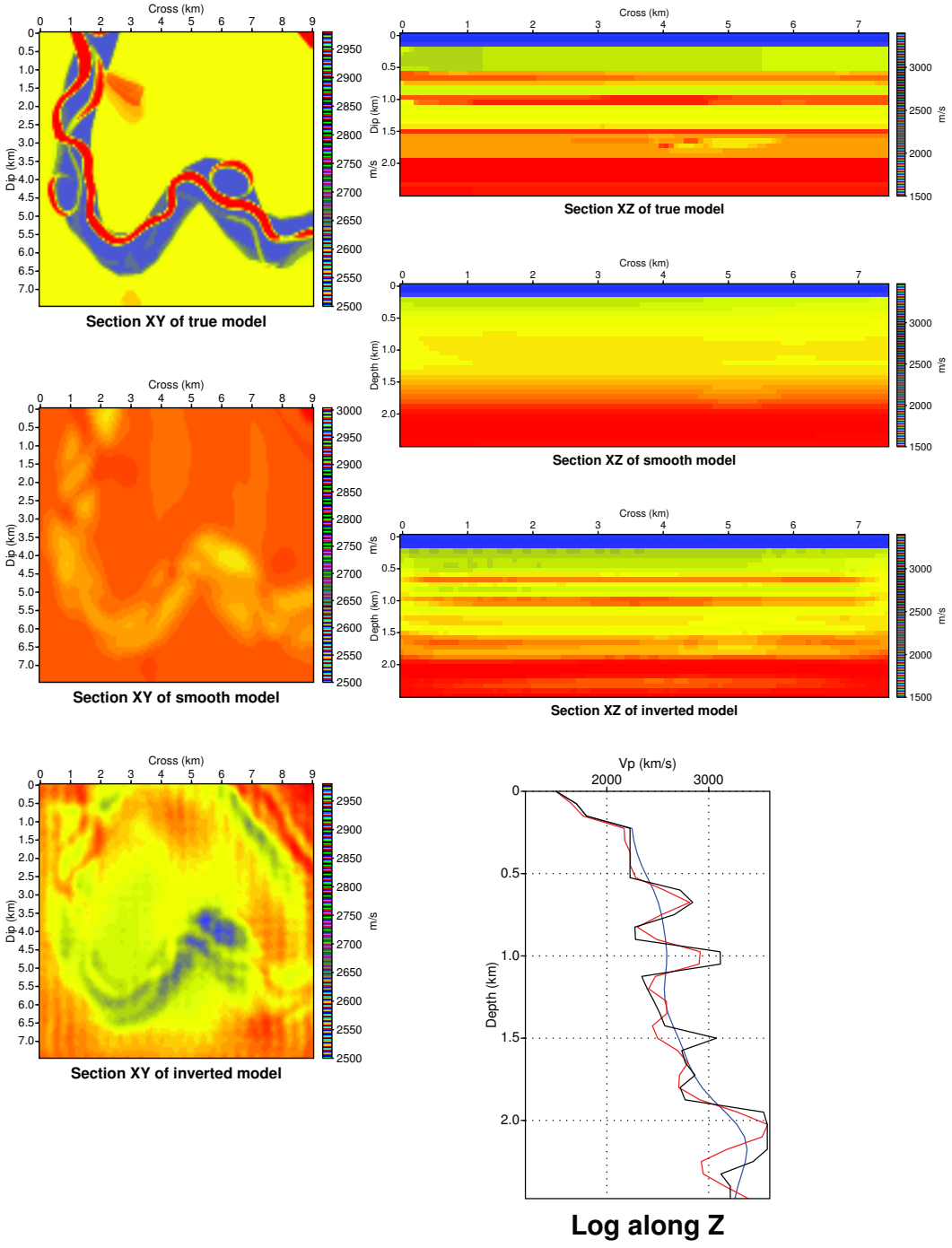

Figure 2.32: Offshore model with free surface - FWI results of the first frequency group inversion. (a, d) vertical (y = 3.75 km) and horizontal (z = 1.65 km) sections extracted from the true model, respectively. (b, e) Same as (a, d) for the smooth initial model. (c, f) Same as (a, d) for the final FWI model. (g) Vertical profile in the middle (x= 5 km, y= 3.75 km) of the model. The black, light gray, and dark gray curves are extracted from the initial model, the true model, and the inverted model, respectively.

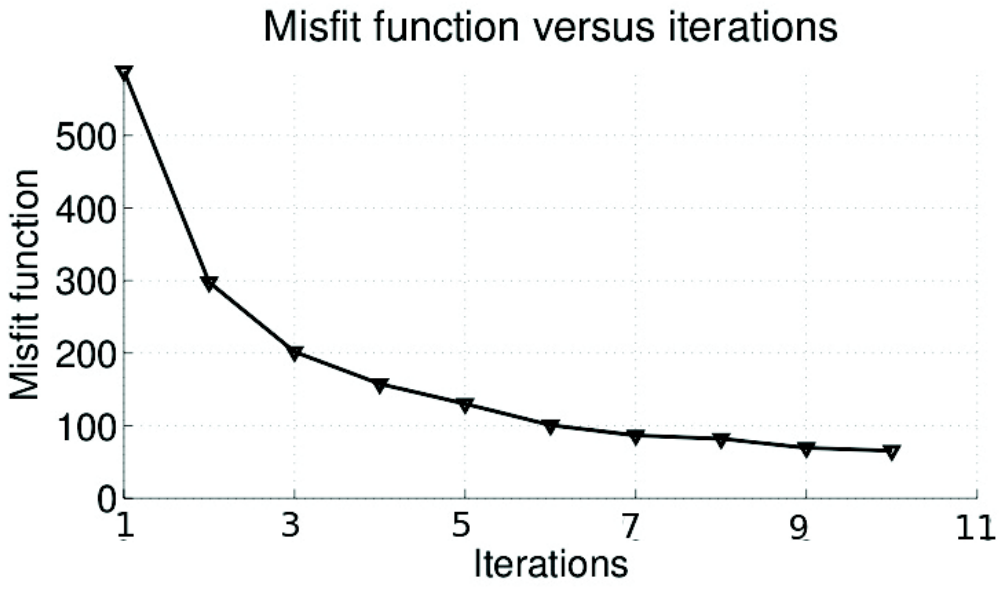

Figure 2.33: Offshore model with free surface - Misfit function versus iterations for the first frequency group inversion. The maximum number of iterations is 10.

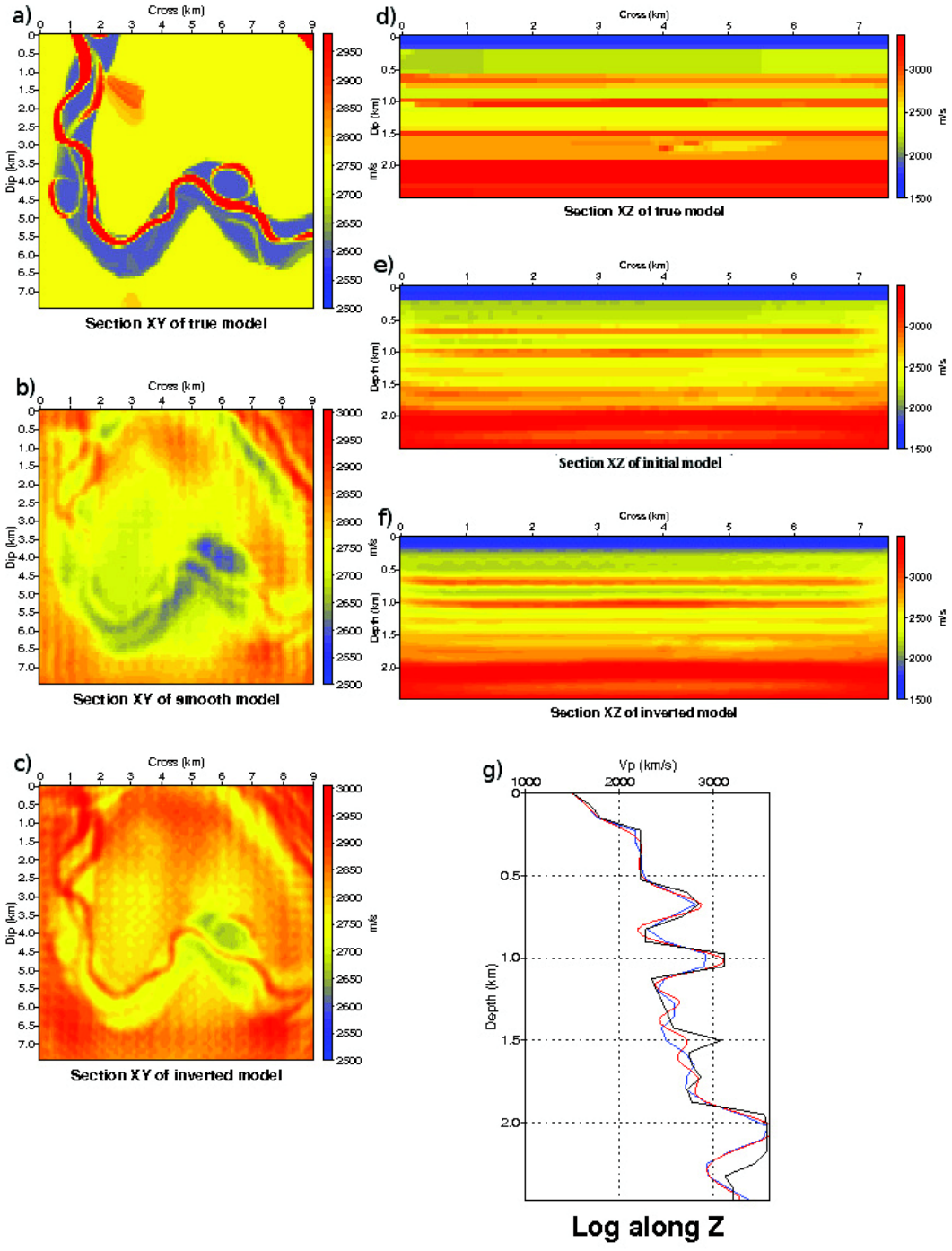

Figure 2.34: Offshore model with free surface - FWI results of the second frequency group inversion. (a, d) vertical (y = 3.75 km) and horizontal (z = 1.65 km) sections extracted from the true model, respectively. (b, e) Same as (a, d) for the intial model. (c, f) Same as (a, d) for the final FWI model. (g) Vertical profile in the middle (x= 5 km, y= 3.75 km) of the model. The blue, black , and red curves are from the initial model, the true model, and the FWI model, respectively.

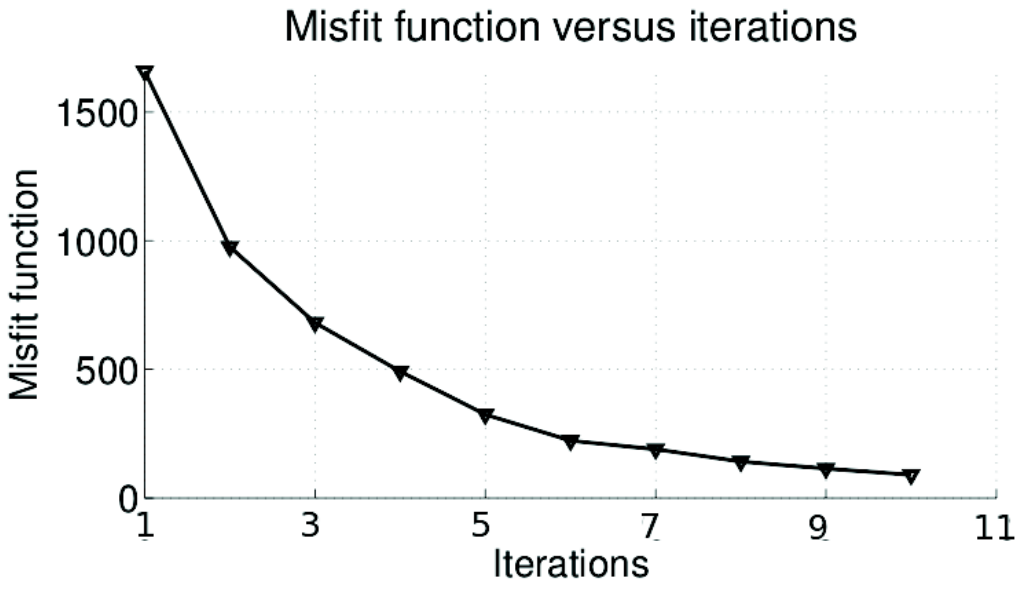

Figure 2.35: Offshore model with free surface - Misfit function versus iterations for the second frequency group inversion. The maximum number of iteration is 10.

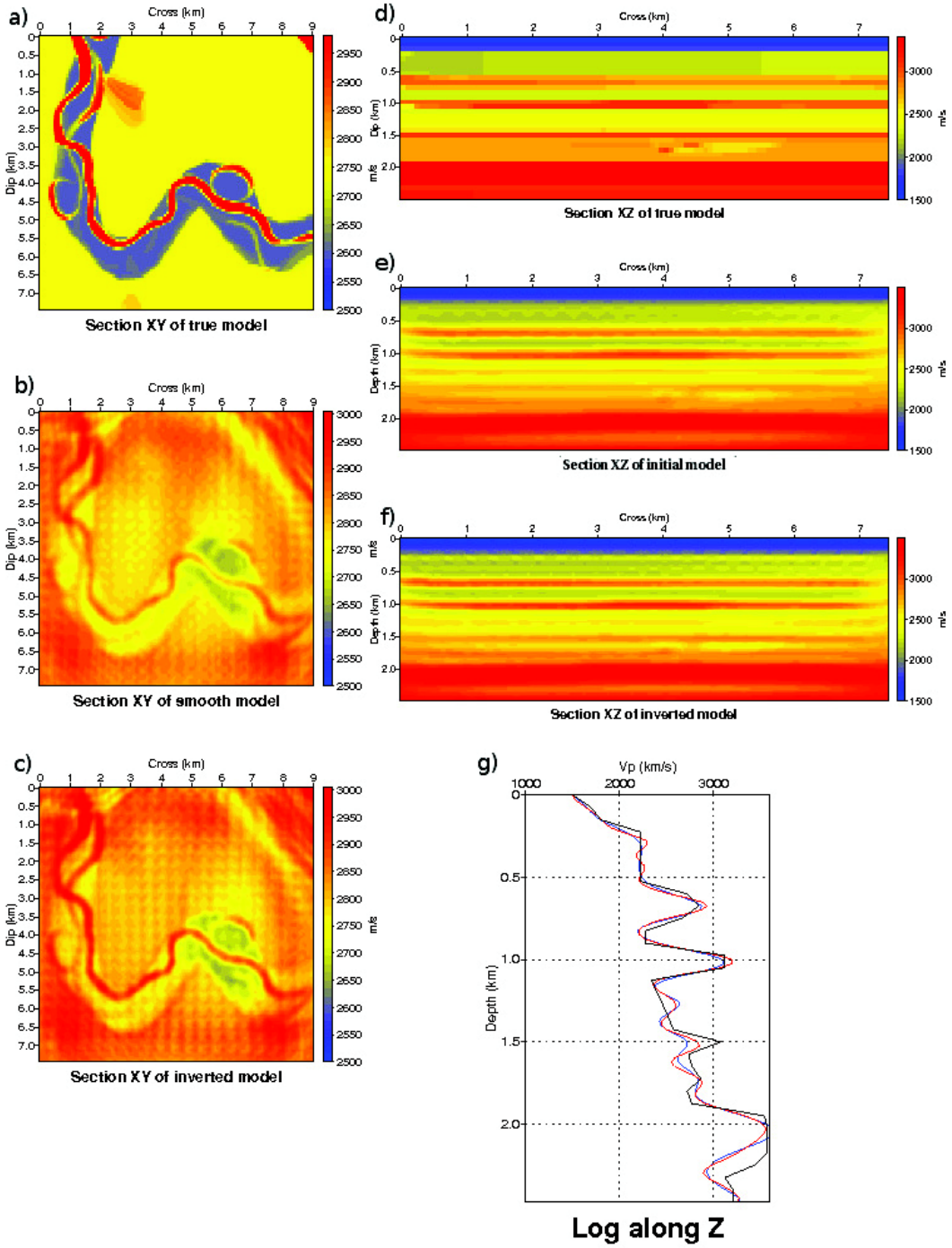

Figure 2.36: Offshore model with free surface - FWI results of the third frequency group inversion. (a, d) vertical (y = 3.75 km) and horizontal (z = 1.65 km) sections extracted from the true model, respectively. (b, e) Same as (a, d) for the initial model. (c, f) Same as (a, d) for the final FWI model. (g) Vertical profile in the middle (x= 5 km, y= 3.75 km) of the model. The blue, black, and red curves are from the initial model, the true model, and the final FWI model, respectively.

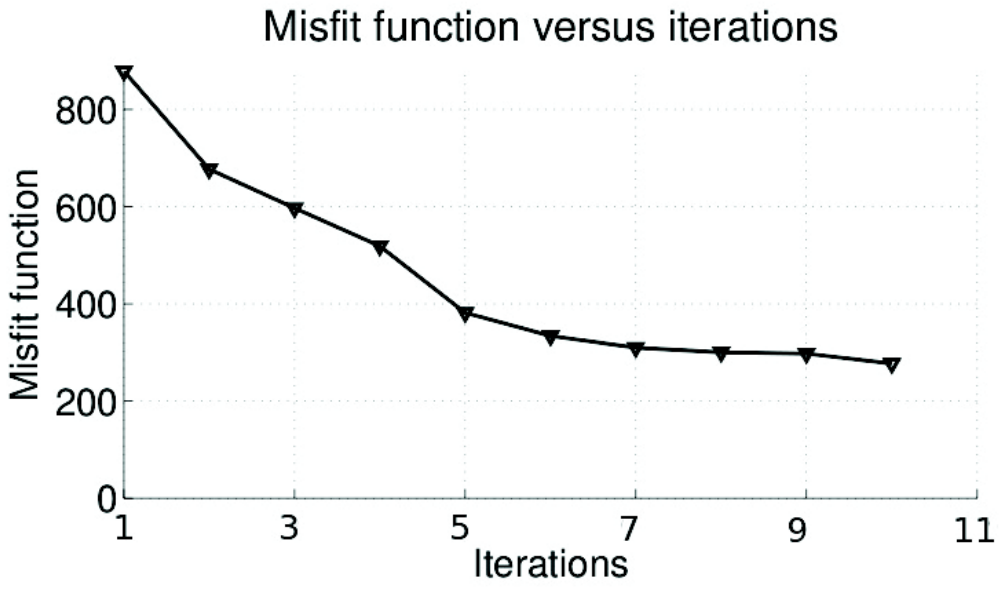

Figure 2.37: Offshore model with free surface - Misfit function versus iterations for the third frequency group inversion. The maximum number of iterations is 10.

2.4 Partial conclusion for the inverse problem

In this chapter, I have introduced the formalism of the inversion in the frequency domain. The inverse problem is solved with a linearized least-squares optimization scheme, which seeks to minimize the misfit between the recorded and predicted seismic data. We have introduced the different line-search algorithms : steepest-descent, conjugate gradient, Quasi-Newton and Gauss-Newton methods. The roles of gradient and Hessian have been presented. They control the expected resolution of the method. The Hessian, as a deconvolution and scaling operator, plays a key role in the convergence and resolution of the inversion.

In the present study, I use a preconditioned conjugate gradient method. I have briefly reviewed the adjoint-state method and I have validated the gradient of the misfit function computed with the adjoint-state formalism against a finite-difference method. I have explained how the adjoint-state method can be applied to a pseudo-conservative form of the wave equation to manipulate selft-adjoint operators, and simplify the scattering kernel of the gradient of the misfit function. This allows us to design a versatile FWI code, where, on the one hand, the same numerical scheme is used to compute the state and adjoint wavefields, and , on the other hand, different modeling engines can be interfaced with the inversion.

A double level of parallelism over shots and domain decomposition has been introduced. I have shown the benefit of combining two parallelisms to tackle large-scale problems, or to save computational ressources. The different validation tests gives us confidence in our work flow. The models obtained by 3D FWI provide high resolution images, which have more potential than the velocity models obtained with other ray-based methods. The geological features can be directly interpreted in the velocity models. These geological horizons can be extracted in a completely automated way from the recovered models rather than from migrated images.

I have simulated a marine case study using a fixed-spread wide-aperture surface acquisition to set up FWI for the real data case study from the Valhall oil field, which is presented in the next chapter.

Part III

Applications

The main aim of this thesis is the application of our FWI algorithm to a ocean-bottom-cable real data set recorded on the Valhall oil field in North Sea. I shall use the FWI strategy, which was validated against the synthetic marine case study in the previous chapter.

In the following of this thesis, I shall discuss two applications. In the first one, I will apply our FWI algorithm to the full overthrust model to tackle a problem of realistic size (full EAGE/SEG Overthrust). In this synthetic test, I will use the so-called inverse crime: the same forward modeling is used to compute the seismic data in the true model and in the models updated during FWI.

The second application is the application of FWI to a real data set from Valhall. Indeed, the real data application is more challenging, because several wave phenomena related to anisotropy, elasticity, and attenuation will not be taken into account during seismic modeling. We have also to deal with potential inaccuracies of the isotropic starting model, which must be sufficient close to the true model to avoid convergence towards a local minimum of the misfit function. Noise in the data can also hamper the convergence of the inversion. In the case of Valhall, we saw that the platform creates a significant amount of noise in the shot gathers located in the vicinity of the platform (Prieux, 2012). The initial model that is used for this application was originally built by vertical transverse isotropic (VTI) reflection traveltime tomography, and converted into normal moveout velocity model for isotropic FWI. The FWI is applied without regularization and data preconditioning. I will first present the Valhall data set, the initial model, and the FWI set-up. Then, I shall show the FWI results, which will be appraised in the final part by seismic modeling, well log analysis, and source wavelet estimation.

Chapter 3

Synthetic and real data applications

Contents

3.1 Synthetic case study : onshore EAGE / SEG Overthrust model

3.1.1 The overthrust model and FWI set-up

We apply our inversion scheme to a synthetic data set computed in the full onshore EAGE / SEG Overthrust model (Aminzadeh *et al.*, 1997). The model is a constant density model with variable acoustic wave speed covering an area of 20 $km \times 20\ km \times 4.65\ km$. It represents an onshore complex thrusted sedimentary succession constructed on top of a basement block. Several faults and channels are present in this model. The minimum and maximum velocities in the model are 2.2 km/s and 6.0 km/s, respectively. We have used this model for the validation of our forward modeling engine in the chapter 1.6.1, (see figure 1.28).

We compute the observed data in the true model with a grid spacing of 75 m. It leads to a finite-difference grid of $267 \times 267 \times 63$ nodes. The surface fixed-spread acquisition, which is

representative of the geometry of the real data case study from Valhall, is shown in figure 3.1: 2,160 sources and 40,000 receivers are considered at a depth of 225 m. Sources are deployed along 24 lines, and each line has 90 sources. The shot interval in the inline direction is 200 m, and the distance between source lines in the cross-line direction is 500 m. A Ricker wavelet of central frequency 5 Hz is used as the common source signal. The receivers are uniformly deployed on the same plane than the sources. The receiver interval is 100 m in the in-line and cross-line directions. The FWI set-up and the computational time are shown in table 3.1. Free surface effects are not taken into account for this application, and therefore an absorbing boundary condition is set on top of the model. Inversion is performed with the conjugate gradient method, and the descent direction is preconditioned with the diagonal terms of the pseudo-Hessian of Shin *et al.* (2001a). This application was performed on an INTEL Xeon(R) CPU E5620 (2.40 GHz) cluster Licallo hosted by Observatoire de la Cote d'Azur.

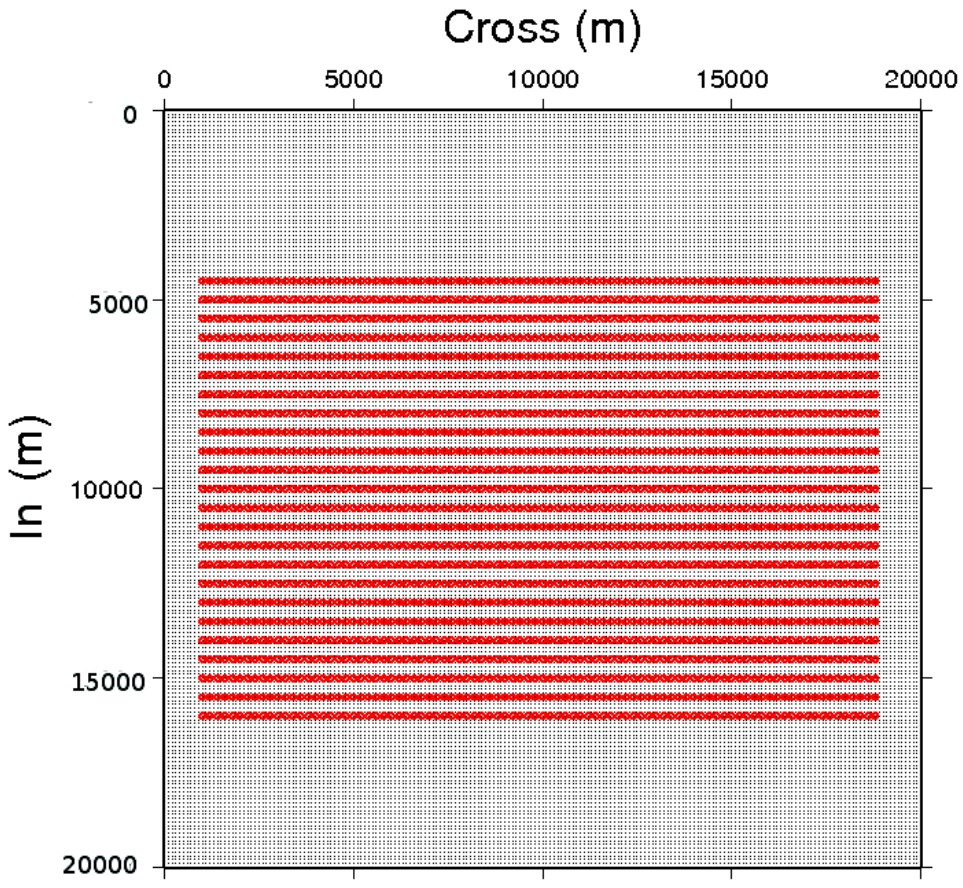

Figure 3.1: Full overthrust model - Fixed-spread surface acquisition. Black points and red lines denote the position of receivers and sources, respectively.

Table 3.1: Full overthrust model - Acquisition geometry and modeling discretization: the dimensions of the model are denoted by N_z, N_x, and N_y in the vertical, eastward and northward directions, respectively; The time step of the forward simulation is denoted by Δt, the total time by T; The grid interval Δh is identical in the three directions; The number of sources is denoted by N_{sour}, the number of receivers by N_{rec}.

N_z	N_x	N_y	Δt (s)	T	Δh (m)	N_{sour}	N_{rec}
63	267	267	0.006	10.	75	2160	40000

3.1.2 FWI results

Following the hierarchical strategy of Pratt *et al.* (1998), we use three slightly-overlapping frequency groups of increasing high-frequency content. The maximum frequency we consider is 7 Hz, taking into account our limited computer resources. A Ricker wavelet of central frequency 2 Hz is used to compute seismic data in the true model. We assume that the source signal is not known, and therefore the source signature is estimated during inversion.

The initial model is obtained by smoothing the true model with a three-dimensional (3D) Gaussian function (the correlation length is 500 m, 500 m, and 250 m in the cross-line, in-line, and depth directions, respectively, as shown in figure 3.2 b and e).

For the first group, we use five frequencies between 3.2 Hz and 4.0 Hz with a frequency interval of 0.2 Hz: [3.2, 3.4, 3.6, 3.8, 4.0] Hz. For this frequency group, the grid interval in the model is 125 m, which satisfies the discretization rule of four nodes per shortest wavelength. For this grid interval, the grid dimensions are $161 \times 161 \times 38$. The source and receiver positions do not coincide with positions of nodes in the finite-difference (**FD**) grid, and therefore the *Hicks* interpolation (Hicks, 2002) is used for source and receiver implementation. The computational statistics and cost of the imaging of the Overthrust model for the first frequency group are outlined in table 3.2. We use 432 processors to perform this test. The parallelism is implemented by source distribution, as we have enough memory per processor (4 GB) to perform single-source modeling. Ten iterations are performed per frequency group. A computational time of 12.5 h is necessary for the inversion of the first frequency group.

The full waveform inversion (FWI) velocity model after the first frequency group inversion is shown in figures 3.2(c, f). On the horizontal slice at the depth of 2.6 km, we show a square pattern superimposed on the structures, besides the blurred channel trace. The size of the square matches the shot and receiver spacing suggesting that it corresponds to the footprint of the coarse acquisition geometry. This footprint of acquisition will be reduced as the inversion progresses towards high frequencies. Figure 3.2g shows a comparison between a vertical profile extracted from true model, the starting model, and the FWI model. The agreement is reasonably good in the vertical direction. Note that the velocity is better estimated at shallow depths in relation with the acquisition geometry. It may be due to an insufficient number of iterations or the low frequency range we use. Indeed, the deep structures are mainly constrained by later-arriving reflections of smaller amplitude. Misfit reduction may be slower for these arrivals since the value of the objective function is dominated by the residuals of the high-amplitude shallow arrivals during the first iterations (Ben Hadj Ali, 2009). The misfit function versus iterations is shown in the figure 3.3. Although the misfit function reduction is low after ten iterations, additional iterations would have allowed to better reconstruct the true

Table 3.2: Full overthrust model case study - FWI set-up and computational time: N_{gre} denotes the inverted frequency group; Dimensions of the model are N_z along the vertical direction, N_x along the East, and N_y along the North; The time step of the forward simulation is denoted by Δt; The grid interval Δh is identical in the three directions; The number of sources is denoted by N_{sour}, the number of receivers by N_{rec}; F (Hz) denotes the frequency band; The number of processors is denoted by N_{core}. N_{ite} is the number of iteration; The computational time for the inversion is denoted by T_{cpu}; The computation is performed in the cluster licallo.

N_{gre}	N_z	N_x	N_y	Δt (s)	Δh (m)	$F(Hz)$	N_{sour}	N_{rec}	N_{core}	N_{ite}	T_{cpu} (h)
First group	38	161	161	0.01	125	3.2–4.0	2160	40000	432	10	12.5
Second group	47	201	201	0.0078	100	4.0–5.0	2160	40000	432	10	32
Third group	63	267	267	0.006	75	5.0–7.0	2160	40000	720	10	52.5

amplitude of velocities.

We successively invert the second frequency group, [4.0, 4.2, 4.4, 4.6, 4.8, 5.0] Hz, and the third frequency group, [5.0, 5.2, 5.4, 5.6, 5.8, 6.0, 6.2, 6.4, 6.6, 6.8, 7.0] Hz. We use the final model of the previous group as the starting model of the next group. The grid interval is adapted to the inverted frequency, that leads to $\Delta h = 100\ m$ and $\Delta h = 75\ m$ for the second and third groups, respectively. The **FD** grid is of dimension $47 \times 201 \times 201$ and $63 \times 267 \times 267$ for the second and third frequency group, respectively. Ricker wavelets of central frequency 3 Hz and 5 Hz are respectively used for the second and third groups. The source signature is estimated. As for the first group, the source and receiver positions do not coincide with positions of nodes of the **FD** grid, and the *Hicks* interpolation is used. The computational statistics and cost of the imaging of the Overthrust model are shown in the table 3.2. We perform ten iterations per frequency group. The computational time is around 32 and 52 hours for the second and third group, respectively. The number of processors is 432 and 720 for the second and third frequency group, respectively.

The FWI velocity models, after the second and third group inversions, are shown in figures 3.4(c, f) and 3.6(c, f), respectively. The corresponding starting models are shown in figures 3.4(b, e) and 3.6(b, e). Comparison between the initial and final FWI models of each frequency group inversion shows a slight improvement of the velocity structure. As for the target model application shown in the previous chapter, the amplitude spectrum of the source wavelet (a Ricker wavelet of central frequency 2 Hz) acts as a data weighting in the misfit function, which penalizes the contribution of the highest frequencies of the group, when multiple frequencies are simultaneously inverted (because all of the frequencies considered for inversion are higher than 2 Hz). It is worth reminding also that the three frequency groups were designed with a one-frequency overlapping. This implies that the contribution of the dominant frequency of a frequency group (i.e., the lowest one according the spectrum of the source wavelet) is already partially present in the starting model, as this frequency was inverted during the previous frequency group inversion.

Figures 3.5 and 3.7 show the convergence rate for the second and third frequency group, respectively. Comparison between a vertical profile extracted from the true, starting, and reconstructed models are shown in figures 3.4g and 3.6g for the second and third frequency group, respectively. The improvement is more obvious in the near surface and in the middle of

the model. As above-mentioned, the deep structures are mainly constrained by later-arriving reflections of smaller amplitude. Misfit reduction may be slower for these arrivals, as the value of the objective function is dominated by the residuals of the high-amplitude shallow arrivals during the first iterations. An amplitude gain with offset applied to the data space may help to speed up the reduction of long offset residuals at the partial expense of the short-offset ones during late iterations (equation (2.49)). However, this strategy requires to check that short-offset residuals are sufficiently reduced to avoid propagating errors associated with inaccurate shallow structures. The misfit functions versus iterations are shown in figures 3.5 and figure 3.7. The misfit function still rapidly decreases at the 10^{th} iteration in the third group inversion, therefore, more iterations would have been needed.

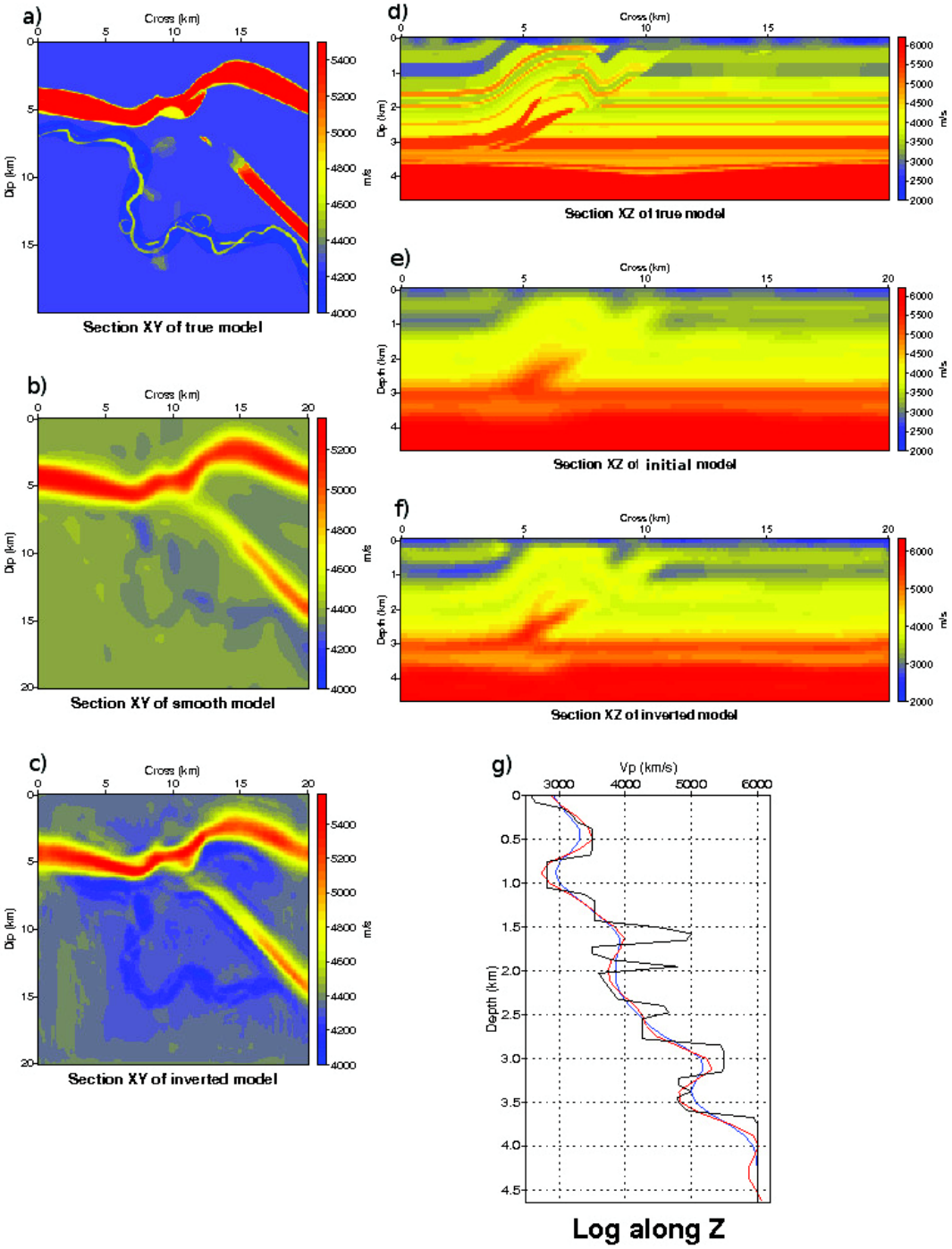

Figure 3.2: Full overthrust model case study - Results of first frequency group inversion. (a, d) vertical (z=2.6 km) and horizontal (y= 5.3 km) sections of the true model, respectively. (b, e) Same as (a, d) for the initial model. (c, f) Same as (a, d) for the FWI model obtained by inversion of the first frequency group inversion. (g) Log along z axis in the middle (x=2.3 km y=5.3 km) of the model.The blue, black and red curves are from the initial model, true model and inverted model, respectively.

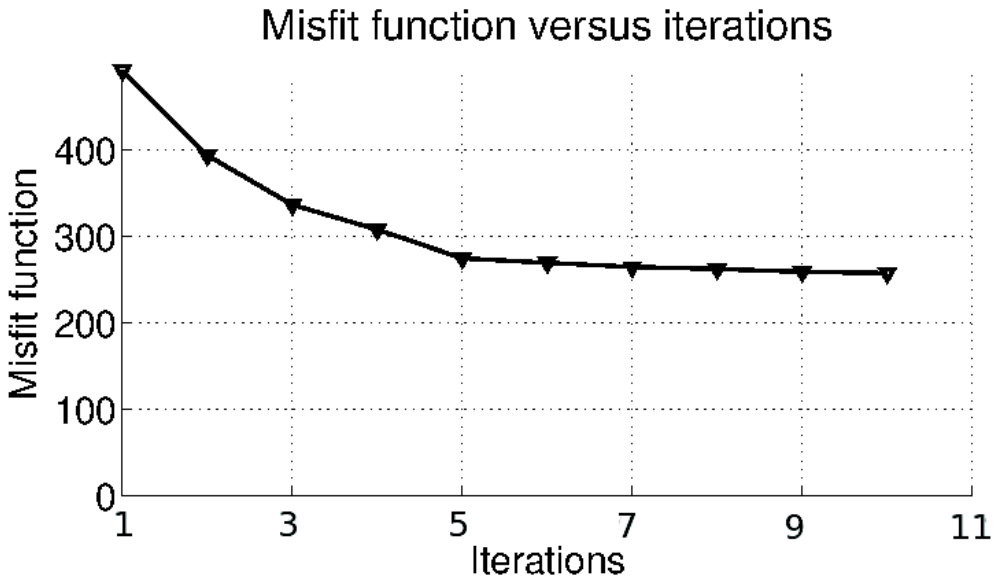

Figure 3.3: Full overthrust model case study - First frequency group inversion. Misfit function versus iterations. The maximum number of iterations was set to 10.

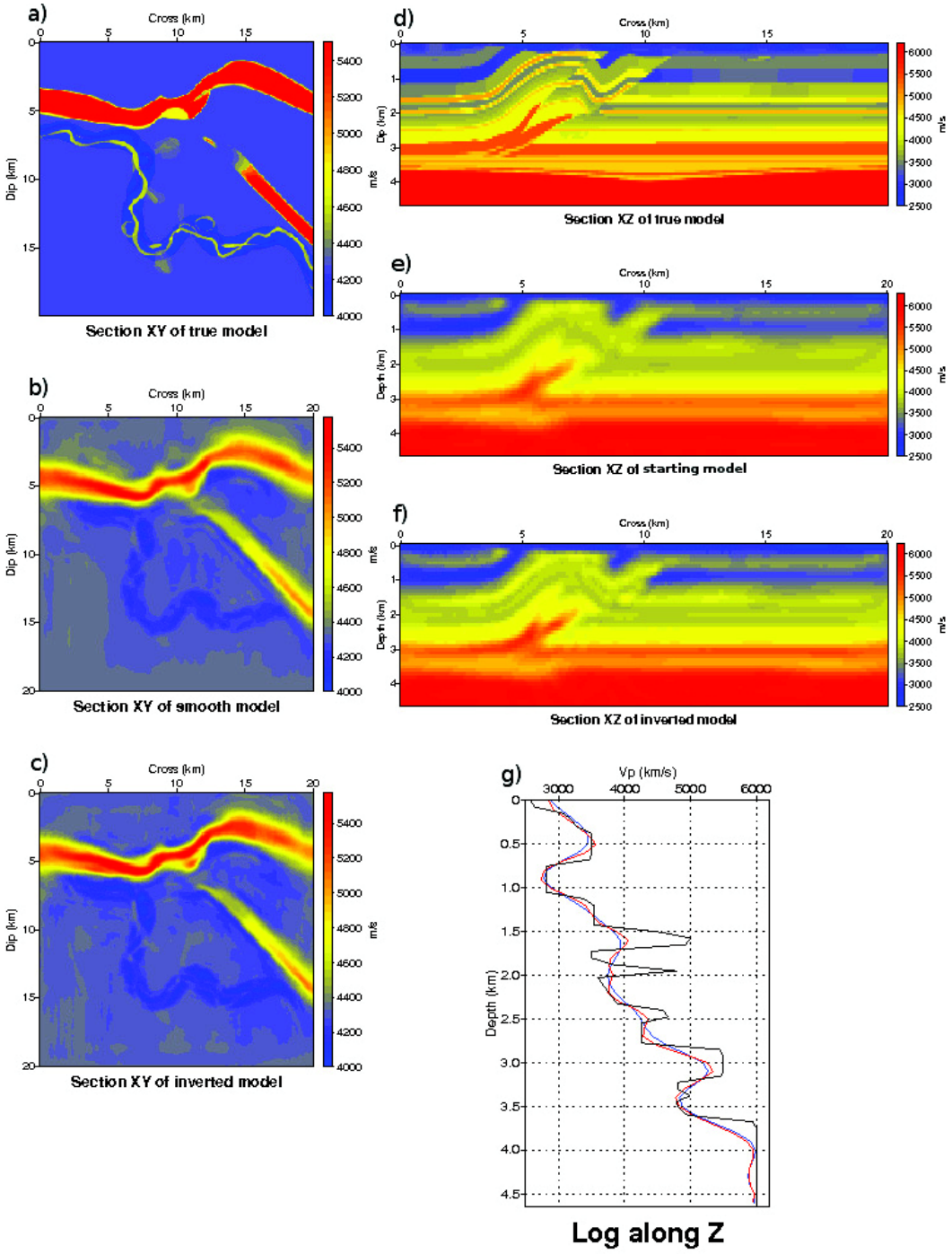

Figure 3.4: Full overthrust model case study - Results of second frequency group inversion. (a, d) vertical (z=2.6 km) and horizontal (y= 5.3 km) sections of the true model, respectively. (b, e) Same as (a, d) for the initial model. (c, f) Same as (a, d) for the final FWI model. (g) Vertical profile extracted in the middle (x=2.3 km y=5.3 km) of the model.The blue, black and red curves are from the initial model, the true model, and the final FWI model, respectively.

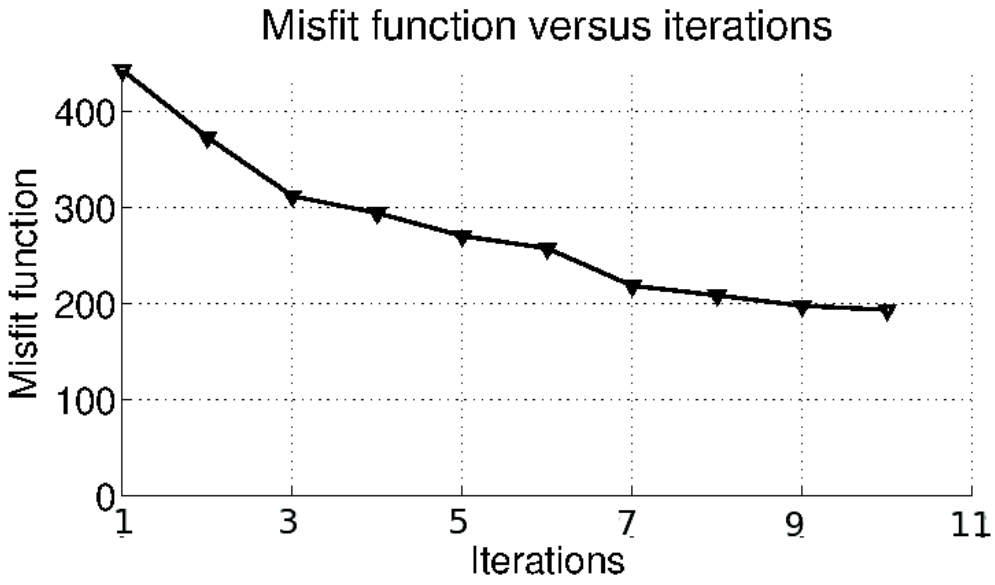

Figure 3.5: Full overthrust model case study - Second frequency group inversion. Misfit function versus iterations. The maximum number of iterations was set to 10.

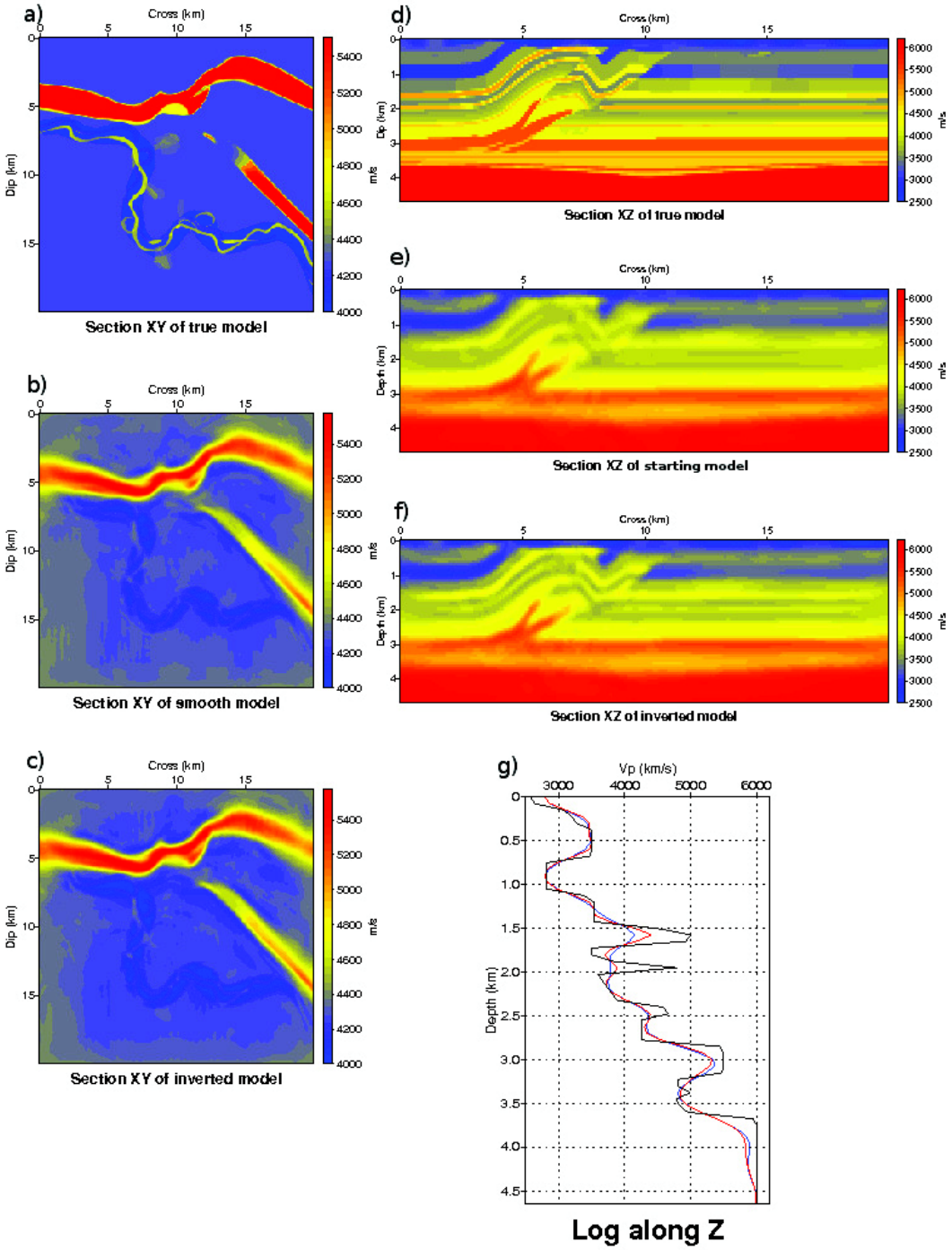

Figure 3.6: Full overthrust model case study - Results of third frequency group inversion. (a, d) vertical (z=2.6 km) and horizontal (y= 5.3 km) sections of the true model, respectively. (b, e) Same as (a, d) for the initial model. (c, f) Same as (a, d) for the final FWI model. (g) Vertical profile extracted in the middle (x=2.3 km y=5.3 km) of the model.The blue, black, and red curves are from the initial model, the true model, and the inverted model, respectively.

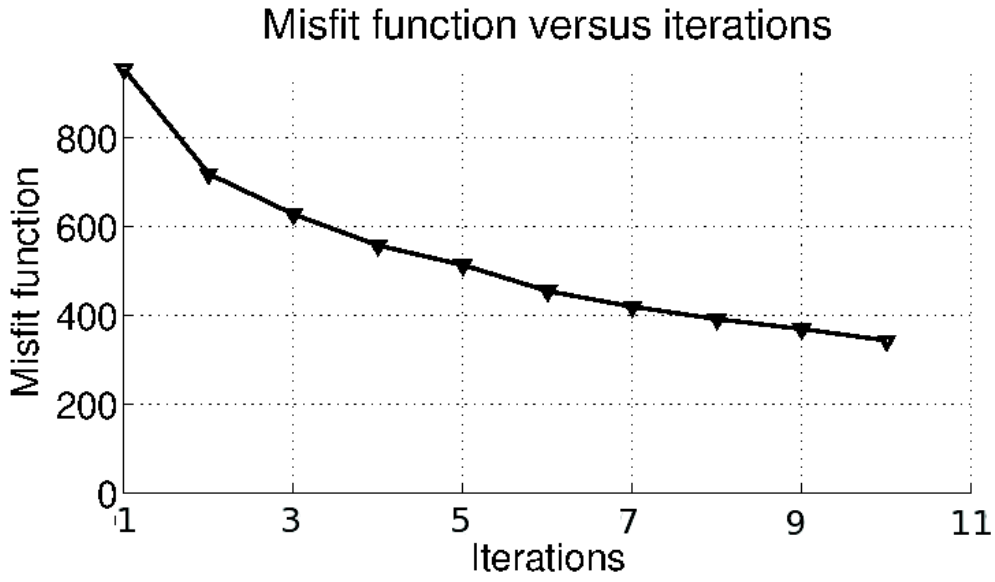

Figure 3.7: Full overthrust model case study - Third frequency group inversion. Misfit function versus iterations. The maximum number of iterations was set to 10.

3.2 Real data case study from the Valhall oil field

3.2.1 Introduction

The Valhall field is located in the Central Graben, North Sea, in the southernmost corner of the Norwegian continental shelf. It was discovered in 1975, and operated by British Petroleum (BP) Norge AS. The field has been in production since 1982, and plans have recently been approved to extend the life of the field until 2048. The main reservoir of the Valhall field consists of a high-porosity and low-permeability embedding rocks. These rocks are weak and collapse under production. Therefore, it is difficult to produce, as initially estimated, 3 billion barrels in place. This compaction results from the subsidence above the reservoir, that affects the entire overburden up to the seafloor, and complicates the drilling of many of the wells that are needed to develop the field. The reservoir is generally complex and there are gas-charged sediments in the overburden, a 'gas cloud', above the top of the field. Seismic imaging has generally been a challenge at Valhall due to the presence of gas in the overburden. Until the last decade, the center of the field was a no-data zone.

Sirgue *et al.* (2009) and Sirgue *et al.* (2010) have been the first to present application of three-dimensional (3D) acoustic full waveform inversion (FWI) to an ocean-bottom-cable (OBC) data recorded on the Valhall field.

In the following of this thesis, I shall present our application of 3D acoustic isotropic FWI to a wide-azimuth data set from the Valhall oil field. The results were obtained with the FWI code *GeoInv3D*, which was described in the previous chapters. I shall analyze the FWI velocity models by means of local comparison of the FWI results with sonic log, source-wavelet estimation, and seismic modeling in time and frequency domain. Although the FWI succeeds in imaging realistic subsurface features, seismic modeling suggests that FWI is hampered by cycling skipping artifacts during the inversion of the early arrival recorded at long offsets.

3.2.2 Geological context

The Valhall field is located in the southern part of the Norwegian North Sea at a water depth of 70 m. The main reservoir of the Valhall field consists of a highly-porous ($42\% \sim 52\%$), low-permeability ($1 \sim 10$ mD), Cretaceous chalk at a depth of about 2400 m. The reservoir thickness varies from 10 m to 60 m, and covers an area of more than 50 km^2. High-porosity reservoir rocks result in significant compaction during pressure depletion. The compaction provides a significant energy drive during production, but also results in subsidence of the overburden that negatively impacts well life and well bore stability. This compaction results in spectacular four-dimensional (4D) seismic responses (Barkved *et al.*, 2009), combined with the overburden unloading. In the time-lapse observations, van Gestel *et al.* (2008) showed the increased pressure and changed saturation associated with production and water injection.

Valhall field was an initially over-pressured, under-saturated Upper Cretaceous chalk reservoir. The reservoir, at a depth of approximately 2400 m sub-sea, consists of two oil-bearing formations: the Tor and Hod. Tor Formation contains the majority of the oil and is a soft chalk characterized by high-porosity.

The stratigraphic log of the flank of anticline is shown in figure 3.8. The lithology, from sea bottom down to the erosion surface of unconformity near 4700 m in depth along with the

geological time varying in age from Permian to Recent, is shown in figure 3.8a. The Central Graben blame into existence during the Triassic, but apparently it had Paleozoic structural antecedents. The oldest sedimentary units of the Greater Ekofisk area comprise Upper Permian Zechstein evaporites, overlain by a non-marine Triassic red-bed sequence and shallow water marine shales. The oldest rocks in Valhall are Upper Jurassic to Lower Cretaceous marine shale. These shales included the Kimmeridge Clay Formation, which is the source rock for the oil at Valhall (Munns, 1985). The first layer composed of interbedding of sandstone and shale, which causes a combination of intrinsic (shale) and layering induced (inter-bedding of shale and sandstone) anisotropy at shallow depth. The second formation is a mixed of siltstone, claystone, and shale lithologies, which has the highest (intrinsic) anisotropy values. Claystone interbedded with limestone is the reason of anisotropy like inter-bedding of sandstone/shale withing fist layer. This formation is the cap rock of below reservoirs. The fracturing is a common phenomenon in limestone formations when confronted with tectonic stresses. The factorization in limestone (due to faults on the crest of anticline) is one main characteristic of this field. Next formations are the reservoirs of Valhall field, the Upper Cretaceous Tor and the lower Hod formations. Tor is an argilious chalk and Hod has chalk/limestone lithology. A cross

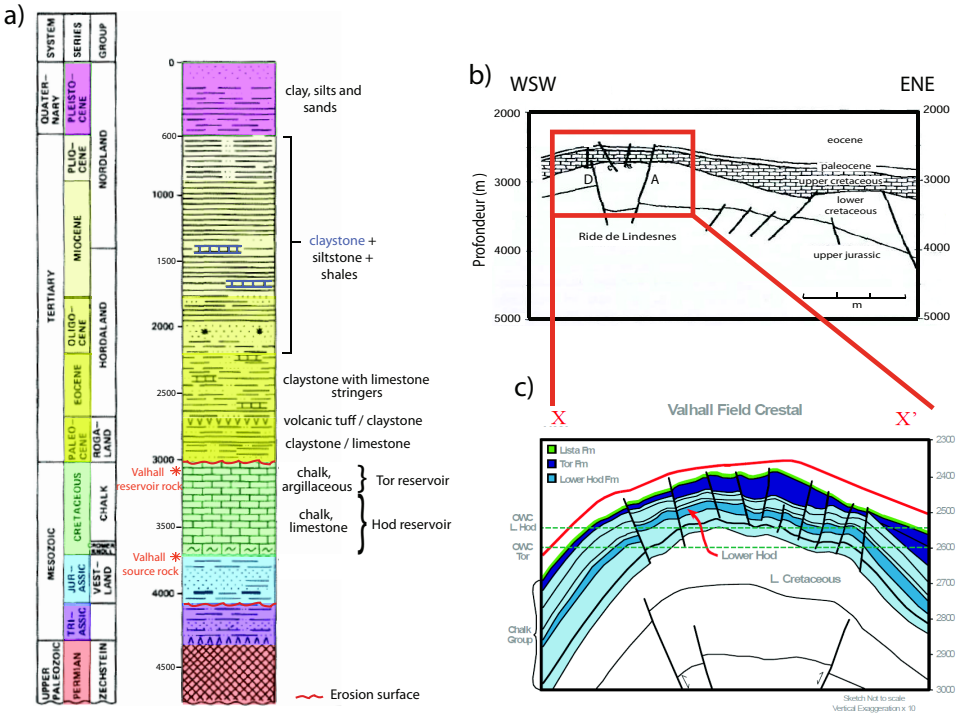

Figure 3.8: Valhall case study - (a) The stratigraphic log of the Valhall field after Munns (1985, figure 5) and Olofsson *et al.* (2003). The log shows the geological formation mostly situated on the flank of the anticline. The cap rock is the Paleocene claystone/limestone formation. (b) A cross section of the anticline (Munns, 1985, figure 20) with graben structure representing the inversion of tectonic stress during geological periods. (c) The close-up from (b) representing the reservoirs composed of Tor and Hod formations. The reservoirs are partitioned by a series of faults on the crest.

section of the field from WSW to ENE direction is shown in figure 3.8b, which demonstrates the inversion of tectonic stresses during geological period. A close-up of the Tor and Hod reservoirs shown in figure 3.8c highlight the partitioning of the reservoir in the anticline crest by many faults, which produce graben structure on the crest. The fractured zone on the crest makes some gas content of reservoir to escape and build the gas cloud. Therefore, the porosity is exceeding 50 percent in places. Some fractured permeability is present, but matrix permeability is generally low, less than 10 mD. The reservoir is an over-pressured, under-saturated reservoir (O. Barkved, 2003). Preservation of such high porosity is primarily due to over-pressured formation, which inhibits mechanical compaction of the chalk (Munns, 1985). The Tor formation is the primary reservoir and the secondary reservoir is from a unit within the Hod Formation. The thickness of Tor reservoir varies abruptly between 0 m and 80 m. Generally, the Tor Formation has higher porosity and permeability, and provides the greatest reservoir volume within structural closure. The Hod reservoir is on average 30 m thick.

3.2.3 Acquisition

The Valhall 3D-4C ocean-bottom cable (OBC) data was primarily designed to improve imaging of the crest of the Valhall structure where conventional seismic imaging is degraded. The primary objective for this reprocessing project was to improve the seismic velocity information in the overburden where the acquisition system and geometry were expected to have a number of effects on the shallow data. The layout of the 3D wide-aperture/azimuth acquisition is shown in figure 3.9. The OBC acquisition design would be expected to provide an ideal data distribution for azimuthal analysis with a wide sampling of both offset and azimuth.

The 2302 receivers were located at around 70 m in depth on the sea floor. The shots were excited at 5m in depth. Each receiver has three orthogonal, nongimbaled geophones and a hydrophone. The distance between receivers is 50m. The minimum distance between the 12 parallel shot cables is 300m. Each receiver records a fine grid of 50000 shots, with a shot interval of 50 m in the dip and cross directions.

3.2.4 Anatomy of data

The data set is composed of 12 cables: cables 11, 13, 15, 17, 19, 21, 23, 25, 27, 29, 31, 35. Each cable contains between 116 and 226 4-component (1 vertical geophone, 2 horizontal geophones, 1 hydrophone) sensors. In this thesis, we used only the hydrophone component, as we focus on acoustic inversion. The data, which were provided to us by BP, were re-sampled with a sampling interval of 0.032 s. The number of samples per trace is 250. Therefore, the trace length is 8 s. The data are sorted by receiver gather. This means that each SEGY file contains the recording of all of the shots by one receiver. One shot profile contains in average 320 sources. The total number of shots is $49,954$. The total number of receivers (that will be processed as shots, in virtue of spatial reciprocity of Green functions) is 2302. The shot spacing is 50 m in the x (cross) and y (dip) directions (Figure 3.9). The receiver spacing along a cable is 50 m. An example of receiver gather data for the shot line, which is co-incident with cable 21, is shown in figure 3.10. We label the different phases in the figure 3.10a: the first-arrival waves (Fa), which mainly propagate in the first 1.5 km of the subsurface (Prieux *et al.*, 2011), the reflections from the top (Rgt) and bottom (Rgb) of gas layers, and the reflections from deeper structures (Rrt, Rrb). We show also a slowly propagating wave, denoted by Rs, which may

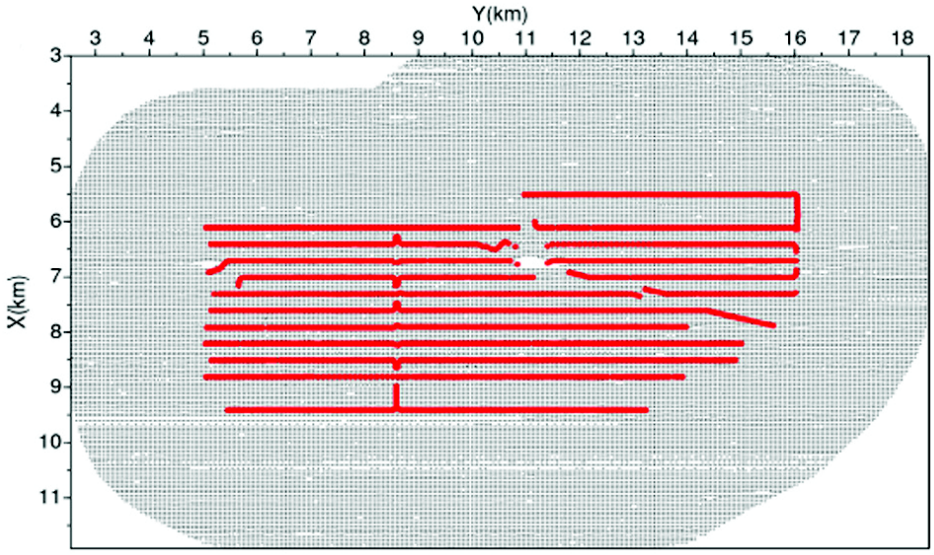

Figure 3.9: Valhall case study - Layout of Valhall survey. Black points and red lines the positions of source and receiver, respectively.

correspond to Stoneley wave. The amplitude spectrum of the data (Figure 3.10c) shows that the frequency bandwidth preserved in the data, after bandpass filtering and under-sampling, ranges between 3 Hz and 14.5 Hz. A monochromatic field at 5.2 Hz is shown in the (x-y) plane in figure 3.11.

3.2.5 The initial model

The three-dimensional (3D) acoustic isotropic velocity model (Figure 3.12) provided by BP was used by Sirgue *et al.* (2009) for 3D FWI of the Valhall data. The initial model (courtesy of BP) was built by vertical transverse isotropic (VTI) reflection traveltime tomography, and was subsequently converted into normal moveout (NMO) velocity and smoothed for isotropic FWI. The dimensions of the initial velocity model in terms of number of grid nodes are $46 \times 91 \times 161$ in the vertical, cross, and dip directions, respectively, for a grid interval of 100 m. Therefore, the physical dimensions of the model are 4.5 x 9.0 km x 16 km. To assess the accuracy of this model, we compute time-domain synthetic seismograms with the wavelet estimated in it (Figure 3.13). Synthetic seismograms computed in the initial model show that the computed first arrivals are delayed compared to the recorded ones, this delay increasing with offset (Figure 3.13, bottom). As the initial model represents NMO velocity, the traveltime mismatch of the first arrivals likely highlights the footprint of anisotropy, as the first arrivals are primarily sensitive to horizontal velocities (Prieux *et al.*, 2011). The time delay reaches a maximum value of 0.24 s, which exceeds half of the period (0.14 s) of the starting frequency that will be used during FWI (3.5 Hz; below 3.5 Hz, the airgun signal was overwhelmed by the background noise). Therefore, the initial model does not allow to satisfy the criterion required to prevent cycle-skipping artifacts during FWI. As the initial model is smooth, the seismic response from the top of the gas layers and from the top of the reservoir are absent from the modeled seismograms. Time delays

between recorded and modeled seismograms translate into significant phase mismatches in the frequency domain. Figure 3.14 shows recorded and modeled 4.06-Hz monochromatic receiver gather in the (x,y) plane. This phase mismatch is significant within the 3$\sim$8 km offset ranges.

We shall see in the Chapter: 3.2.8 how cycle skipping artifacts during FWI translate in the fit between recorded and modeled data in the time domain.

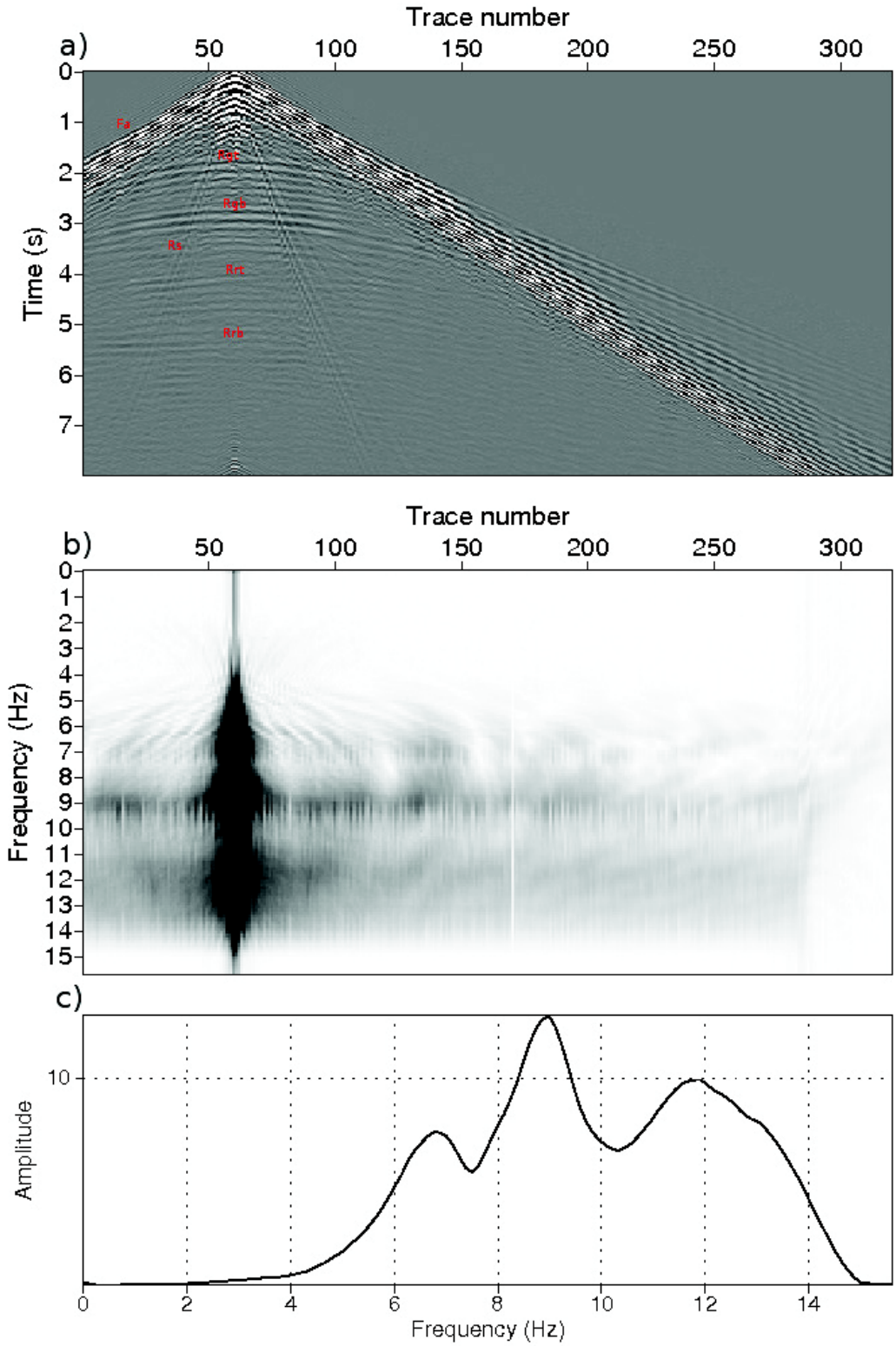

Figure 3.10: Valhall case study - OBC data set. (a) An example of recorded receiver gather of cable 21. The coordinates of the receiver are $x = 6824$ m $y = 15533$ m. The shots are located above cable 21 (b) Spectral amplitudes of the receiver gather. (c) Mean amplitude spectrum of the receiver gather.

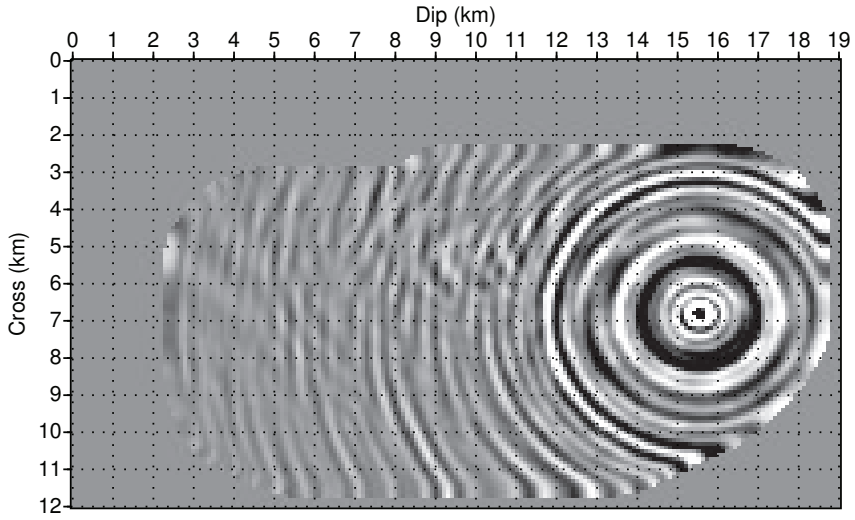

Figure 3.11: Valhall case study - An example of recorded receiver gather of cable 21 after Fourier transform. The real part of the complex-valued monochromatic wavefield is plotted for all of the shots in the horizontal plane defined by the shot positions. The coordinates of the receiver gather are $x = 6824$ m $y = 15533$ m. Frequency is 5.2 Hz.

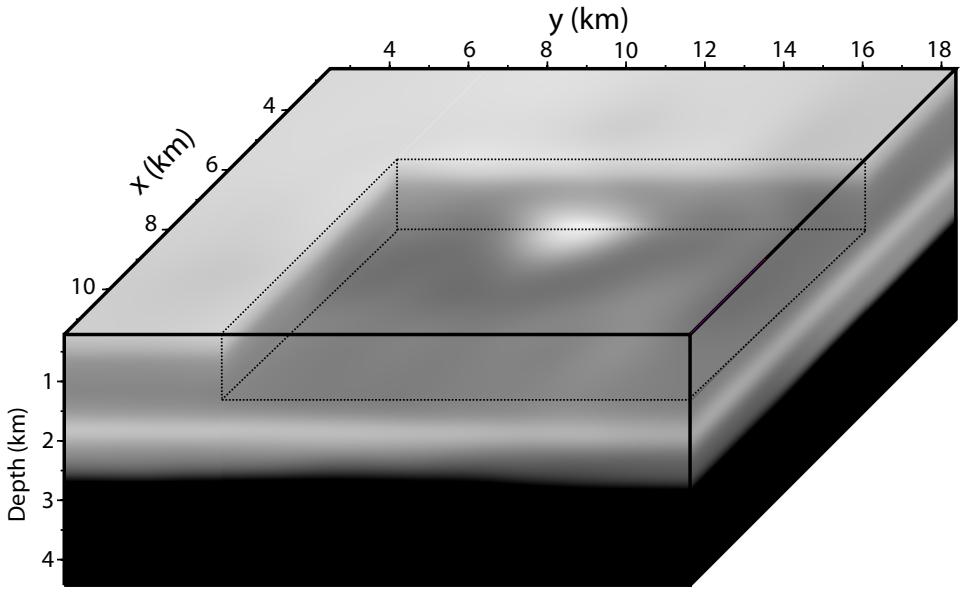

Figure 3.12: Valhall case study - 3D view of the smooth initial velocity model provided by BP. A low velocity zone associated with the gas cloud is visible in the horizontal plane at around 1 km in depth. However, the resolution is quite low. The minimum and maximum velocities of the black & white scale are 1.6 km/s and 2.7 km/s, respectively.

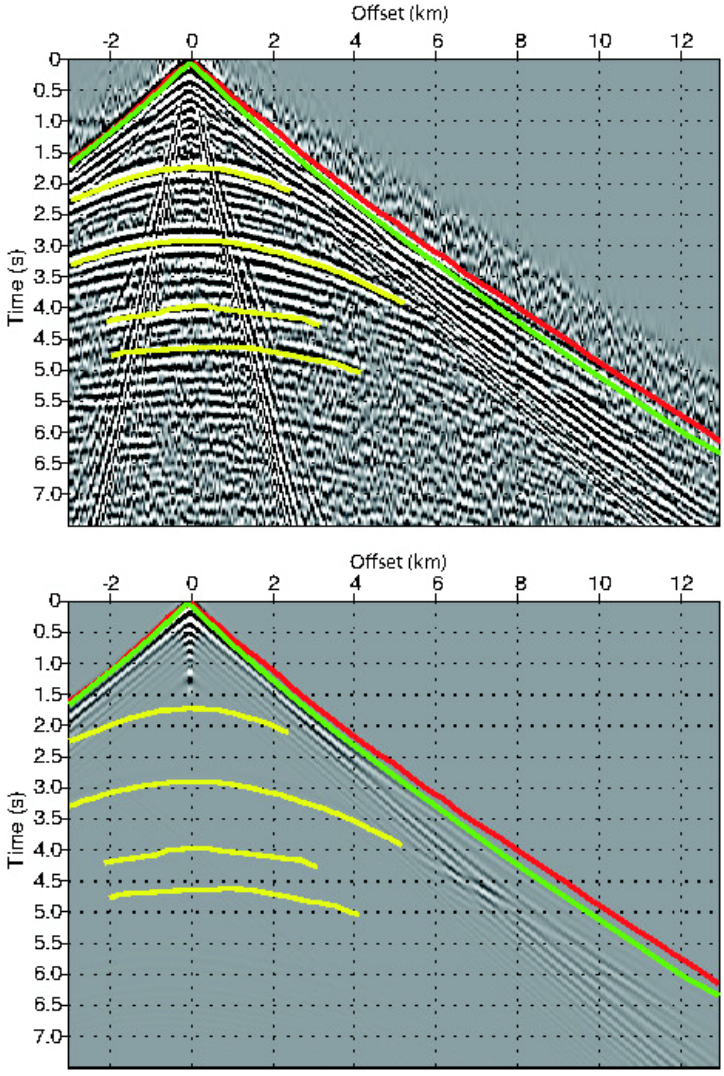

Figure 3.13: Valhall case study - Top: An example of receiver gather. Picks of the main arrivals are superimposed. Bottom: the same receiver gather computed in the initial model. The green line corresponds to the first-arrival traveltimes computed in the initial model; the red line corresponds to first-arrival traveltimes picked in the recorded data; the yellow lines correspond to reflection picks from top and bottom of the gas cloud and the deeper structures.

3.2.6 FWI data preprocessing and experimental setup

We apply FWI to the hydrophone component of the Valhall dataset provided by BP. This dataset was already preprocessed for FWI: the data were undersampled with a sampling rate of 32 ms, and bandpass filtered such that the frequency components greater than 15 Hz are zero (Figure 3.10). A mute was also applied above the first-arrival time (Figure 3.13).

A first quality control of the data in the frequency domain was performed at a frequency of 4.06 Hz, by scanning all of the receiver gathers in the (x,y) plane. Monochromatic data are computed in the initial model, and source estimation is performed for this frequency. Comparison between recorded and modeled data at near offset is use to validate our source estimation procedure in the FWI code (Figure 3.14). We identify several shots that have a poor signal-to-noise ratio. These shots are located near the platform (Figure 3.15). They do not bring any significant information for FWI. The residuals between the recorded and computed data are much more important than the synthetic data, and nearly identical to the observed data (Figure 3.15d). Therefore, around twenty noisy shots were removed from the data set. We successively applied our FWI algorithm to three overlapping frequency groups between 3.5 Hz and 7.0 Hz: the first frequency group contains five frequencies, [3.56, 3.69, 3.81, 3.93, 4.06] Hz, the second one contains 9 frequencies, [4.06, 4.18, 4.31, 4.42, 4.55, 4.67, 4.79, 4.92, 5.04] Hz, and the third one contains four frequencies, [5.04, 5.78, 6.52, 7.01] Hz. Note that there is one-frequency overlapping between two next frequency groups. The starting frequency (3.5 Hz) is the same than the one used by Sirgue *et al.* (2010). The maximum frequency is 7.0 Hz, which requires a grid interval of 50 m for seismic modeling. Considering higher frequencies for FWI would have required more important computational resources than the available ones.

The water depth is only 70 m in Valhall, and as a result, it is a a relatively noisy environment. At progressively low frequencies, noise from the field operations and from the ocean waves dominates, as the airgun signals fade away (Dellinger et Yu, 2009). Only the hydrophone component data is used for acoustic FWI. The inversion is applied without any regularization and data preconditioning. The conjugate gradient algorithm is used for optimization. The descent direction is preconditioned by the diagonal terms of the pseudo-Hessian to better balance the shallow and the deep model perturbations. The source signature is estimated for each receiver gather (in virtue of source reciprocity) during each non-linear iteration of the FWI by solving a linear inverse problem (Pratt, 1999). The initial model (courtesy of BP) is built by a VTI reflection traveltime tomography, and is subsequently converted in NMO velocity and smoothed for performing isotropic FWI (Figure 3.12). Seismic modeling is performed with a $\mathcal{O}(\Delta x^4, \Delta t^2)$ finite-difference method, which requires 5 grid points per wavelength. The grid interval h is set to 70 m for the first two frequency groups such that the finite-difference grid matches the sea bottom, and is refined to 50 m for the last frequency group inversion to prevent numerical dispersion during seismic modeling. The density is assumed to be constant, and is set to 1000 $kg.m^{-3}$. During modeling, the free-surface effects are taken into account, and therefore multiples are involved in the inversion procedure. The inversion is performed on an IBM Blue Gene (850 MHz) cluster Babel from the IDRIS computational center. The computing statistics are shown in the table 3.2.6, and shows the rapid increase of computational time as the size of the problem increases. One level of parallelism by distribution of shots over processors is used for the first frequency group, because single-source modeling and storage of the four frequency components involved in the inversion can be performed on one core of the IBM Blue Gene. Two patchs of 1536 and 704 shots are processed sequentially on 1536 cores. The two levels

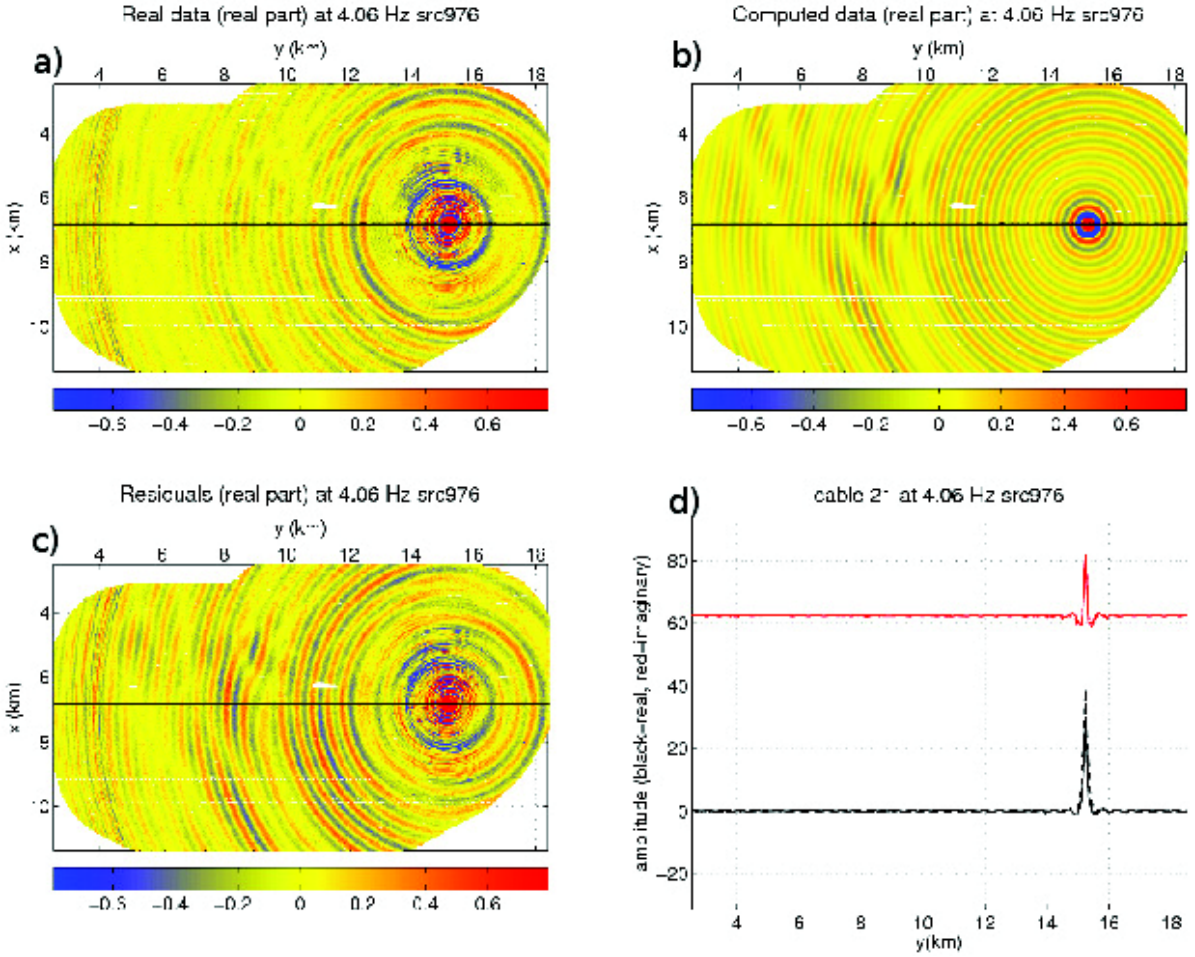

Figure 3.14: Valhall case study - (a) Recorded monochromatic data (real part) at 4.06 Hz. The source, which corresponds to a receiver position, is located at x=6.8 km and y=15 km; (b) Synthetic data computed in the initial model; (c) Difference between (a) and (b); (d) Horizontal profile extracted from the recorded (solid line) and the modeled (dashed line) wavefields. Position of the profile is denoted by the black line in (a). The black and red lines correspond to the real and imaginary part of the complex valued wavefields, respectively. A good agreement is shown at nearest offset, that was used as a criterion to validate the source estimation performed during FWI.

of parallelism are used for the last two frequency groups on 2048 cores. Two subdomains are used, and two patches of shots are processed sequentially.

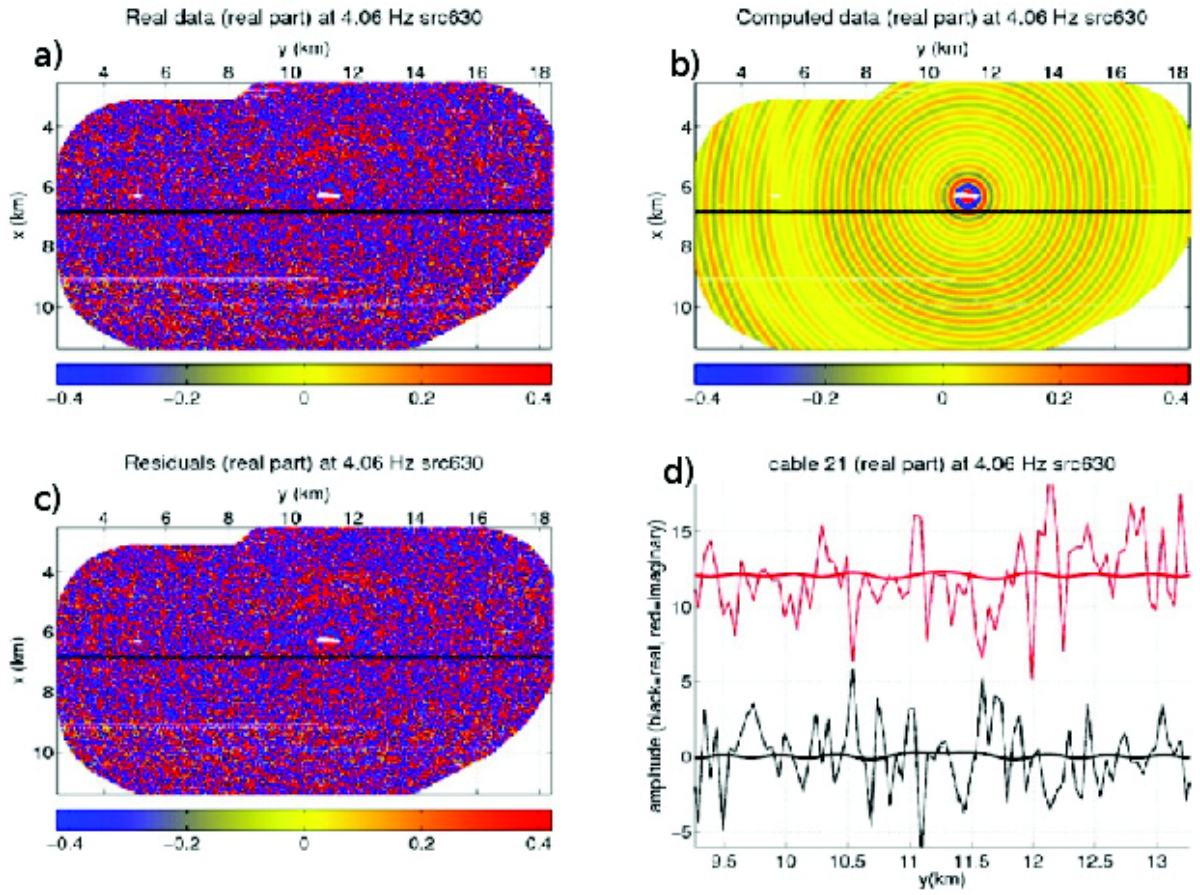

Figure 3.15: Valhall case study - Example of noisy receiver gather: (a) Recorded monochromatic data (real part) at 4.06 Hz. (b) Synthetic data computed in the initial model; (c) Difference between (a) and (b); (d) Horizontal profile extracted from the recorded (solid line) and the modeled (dashed line) wavefields. Position of the profile is denoted by the black line in (a). The black and red lines correspond to the real and imaginary part of the complex valued wavefields, respectively. The residuals are superimposed on the computed data.

Freq. gr.	Max. freq.	h(m)	Nb cores	Nb iter	Comp. time
Gr 1	4 Hz	70	1 536	8	18.0 h
Gr 2	5 Hz	70	2 048	13	29.9 h
Gr 3	7 Hz	50	2 048	6	34.8 h

Table 3.3: Valhall case study - FWI statistics for the three frequency-group inversions: frequency group are denoted by $Freq.gr$; The maximum frequency in the group is denoted by $Max.freq.$; The grid interval h is identical in the three dimensions. The number of processor is denoted by $Nbcores$; The number of iteration is noted by $Nbiter$; The computational time is denoted by $Comp.time$.

3.2.7 FWI results

The final FWI model obtained after successive inversion of the three frequency groups (Figure 3.12) is shown in figure 3.16. It shows quite similar features than those of the model of Sirgue *et al.* (2010): in the near surface (150 m in depth), a complex network of channels is recovered, while, at a depth of 1260 m, the gas cloud is much better described than in the initial model. We observe that FWI models are dramatically improved in resolution, as we proceed over frequency groups. Other perspective views of the final model are shown in figures 3.17, 3.18, 3.19, and 3.20. The main structures that are highlighted in these figures are the shallow channels (Figure 3.17), the horizontal section of the gas cloud (Figure 3.18), a vertical section of the gas cloud as well as a horizontal section across the reservoir (Figure 3.19), and a vertical gas filled fracture (Figure 3.20).

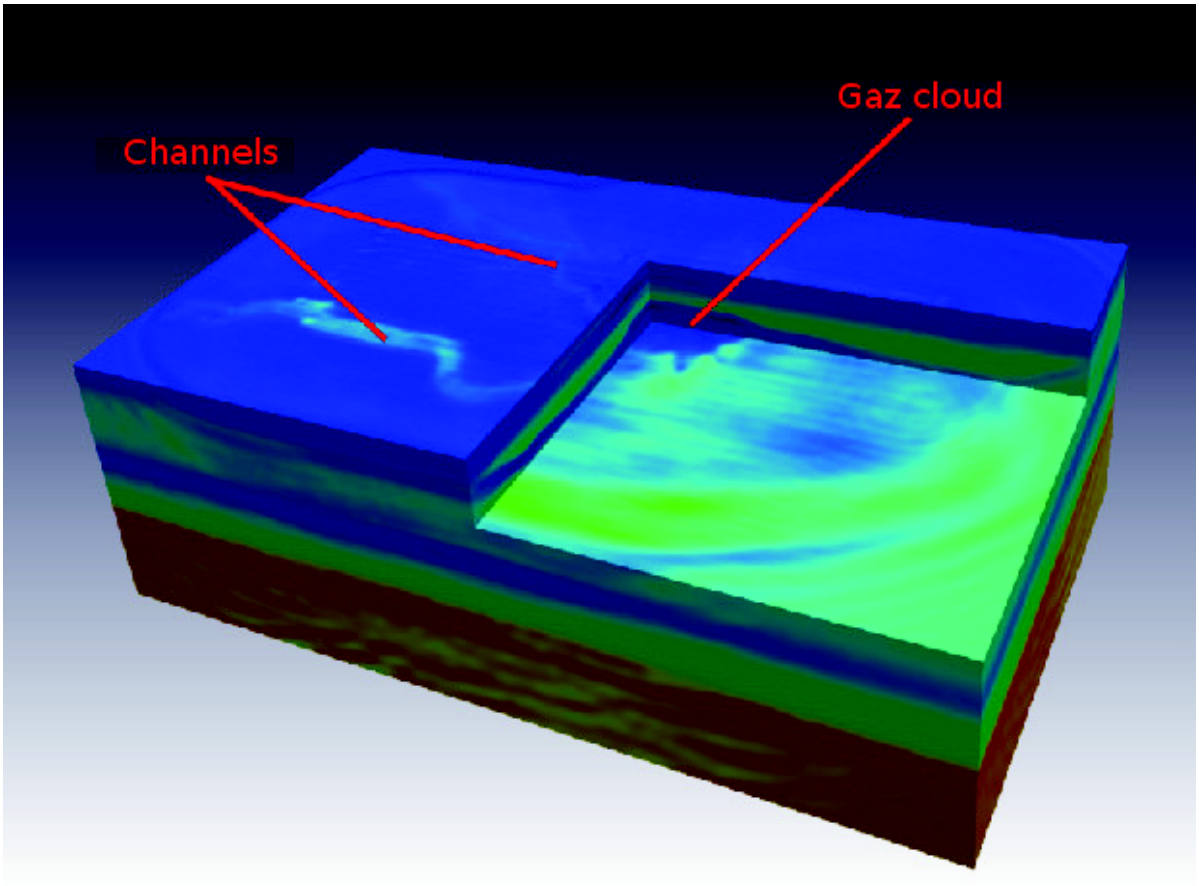

Figure 3.16: Valhall case study - Final FWI velocity model after successive inversions of the three frequency groups: channels (200 m in depth) are built up and the gas cloud (1200 m in depth) is identified.

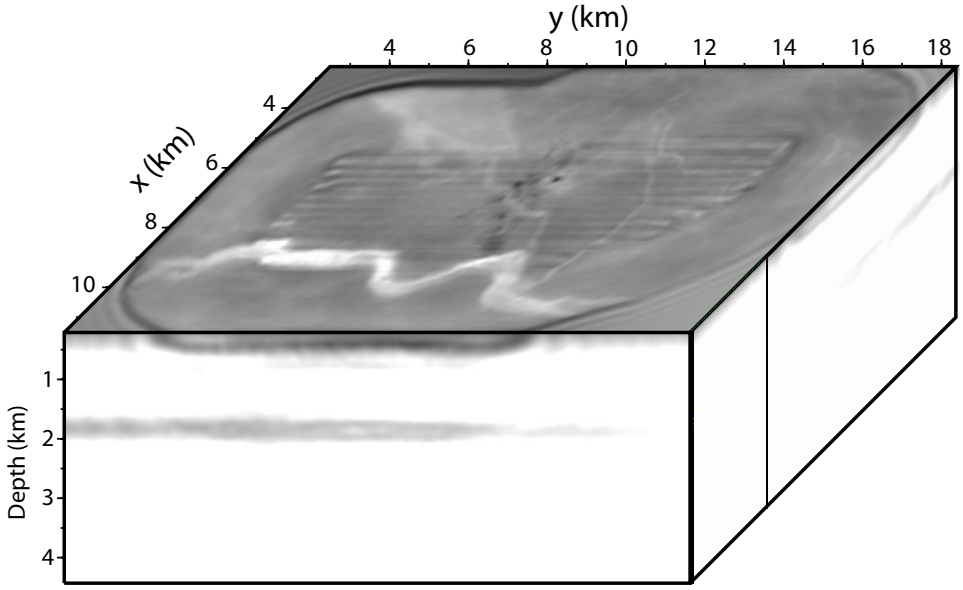

Figure 3.17: Valhall case study - Final FWI velocity model after successive inversions of the three frequency groups: The top section is across channels. The minimum and maximum velocities of the black & white scale are 1.6 km/s and 2 km/s, respectively.

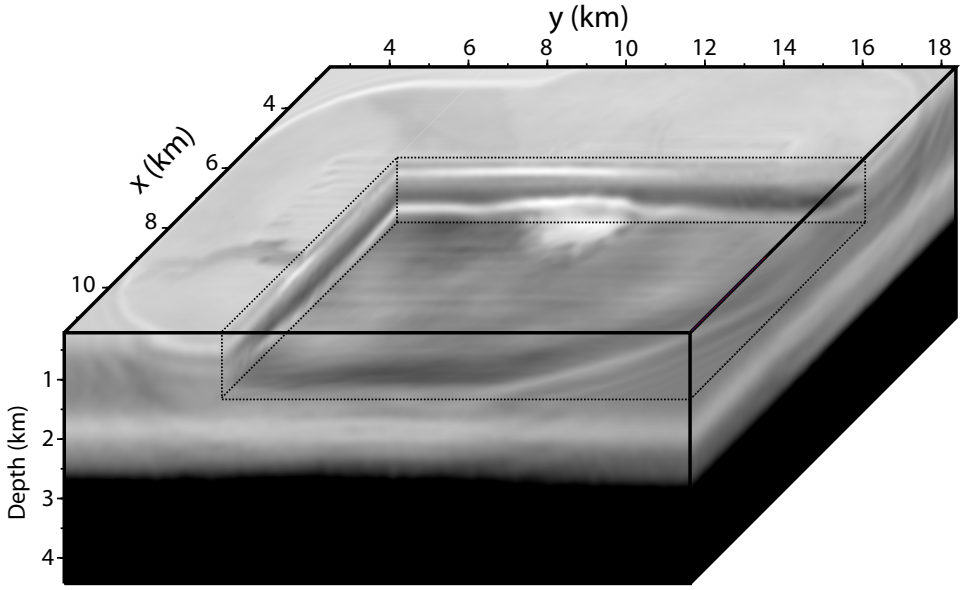

Figure 3.18: Valhall case study - Final FWI velocity model after successive inversions of the three frequency groups: the horizontal section of the gas cloud is highlighted. The minimum and maximum velocities of the black & white scale are 1.63 km/s and 2.73 km/s, respectively.

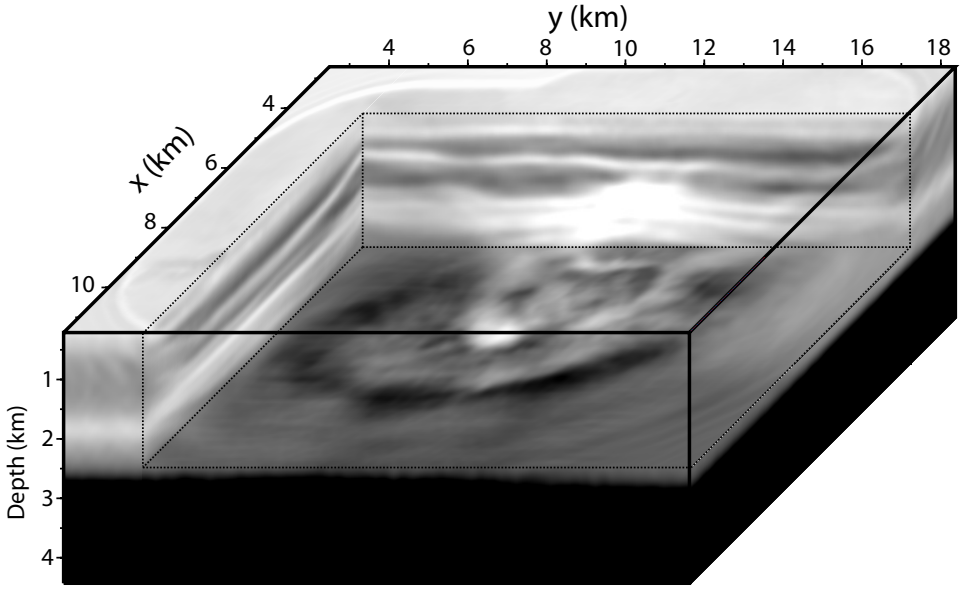

Figure 3.19: Valhall case study - Final FWI velocity model after successive inversions of the three frequency groups: vertical section of the gas cloud as well as a horizontal section at the reservoir level are highlighted. The minimum and maximum velocities of the black & white scale are 1.73 km/s and 2.73 km/s, respectively.

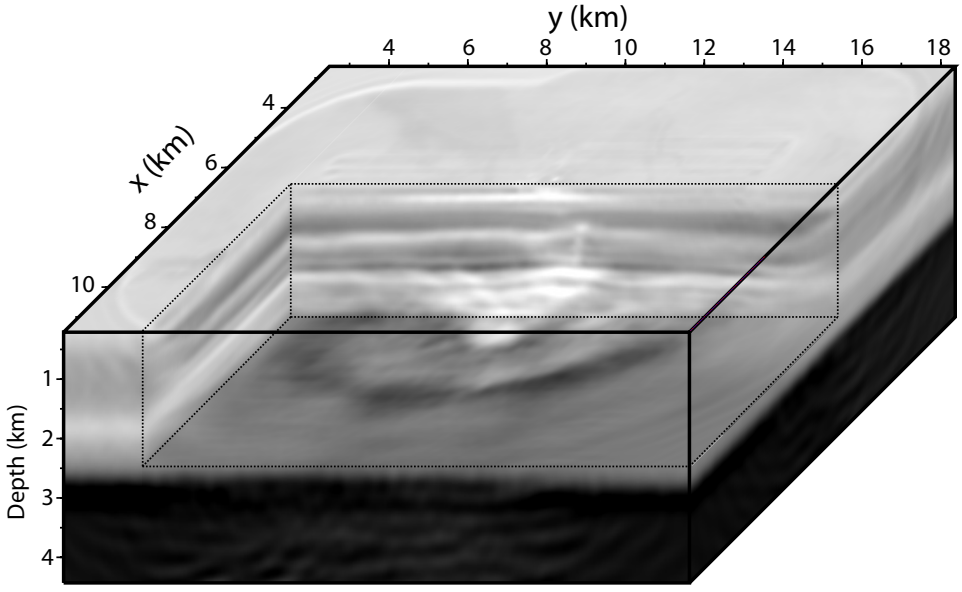

Figure 3.20: Valhall case study - Final FWI velocity model after successive inversions of the three frequency groups: a near-vertical gas filled fracture at y = 12 km is highlighted. The minimum and maximum velocities of the black & white scale are 1.63 km/s and 3 km/s, respectively.

In the following, I will use the following nomenclature to denote the FWI models: the notation mi denotes the FWI model obtained at the total iteration number i. A comparison between a horizontal and vertical slices extracted from the initial model and the final FWI model of the first frequency group($m8$) is shown in figures 3.21 and 3.22 , respectively. The horizontal slice is extracted at a depth of 210 m and 1260 m, respectively (figure 3.21). These horizontal slices cross-cut a complex network of channels and a gas cloud, respectively, which were already imaged by Sirgue *et al.* (2010). The channels, which are absent in the initial model, are well resolved by FWI. However, their image is hampered by the footprint of the cables and to a lesser extent of the shots. At 1260 m in depth, the image of a gas cloud was nicely refined. Although low frequencies are inverted, fine-scale structures, such as gas-filled fractures are clearly identified at the ends of the gas cloud. The acquisition footprint is still visible at 1260 m in depth, although it is muck weaker than in the shallow part of the model.

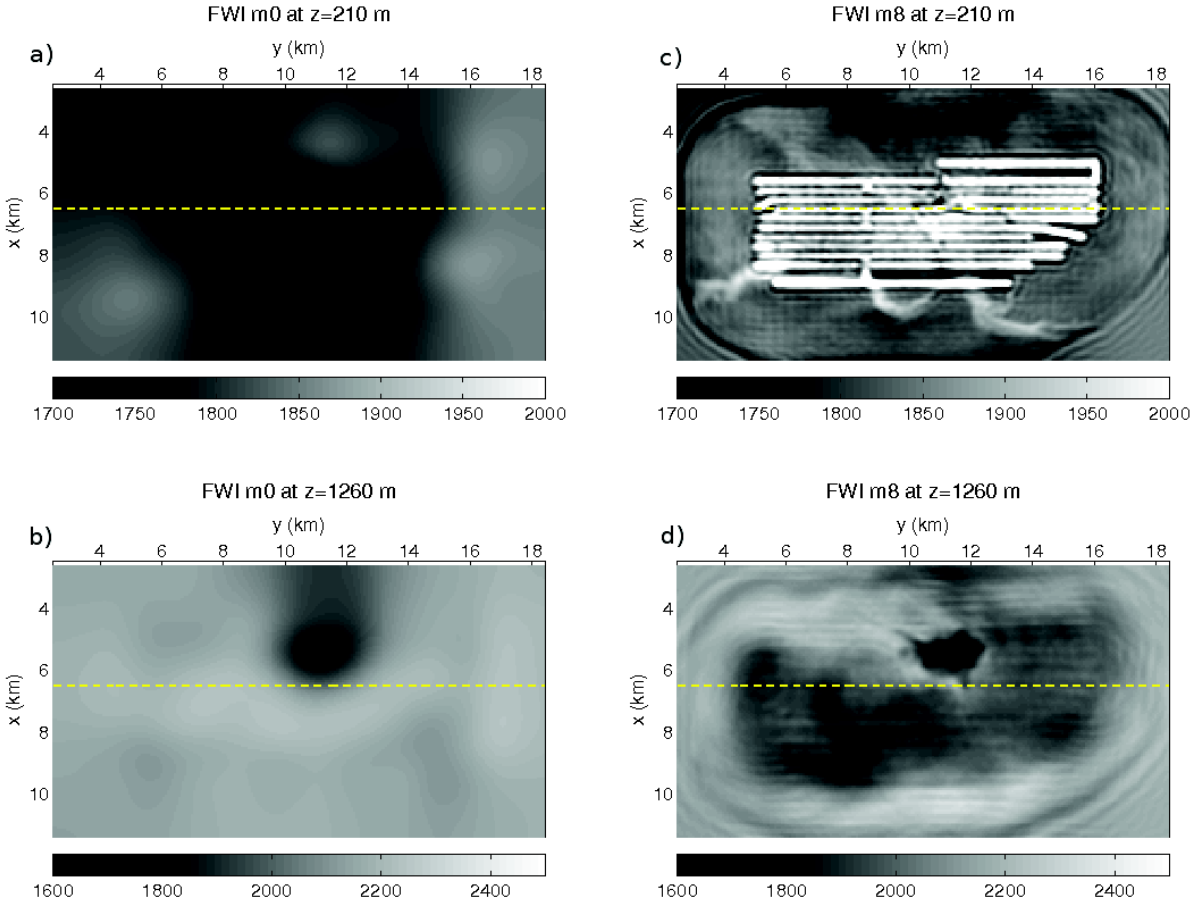

Figure 3.21: Valhall case study - First frequency group inversion: horizontal slice at a depth of 150 m extracted from the initial model (a), and for the final FWI model of the first frequency group inversion (c). Horizontal slices at a depth of 1260 m extracted from the initial model (b) and from the final FWI model of the first frequency group inversion (d). The yellow line denotes the position of an available well log. The channels and the gas cloud, which are barely visible in the initial model, are clearly identified.

The vertical section of the FWI model in the y direction at x=6.5 km shows an interface at 0.6 km in depth, which was already imaged by Sirgue *et al.* (2010) and Prieux *et al.* (2011). Gas layers between 1.7 and 2.7 km in depth were refined after FWI. We note a low-velocity layer at around 1 km in depth, which is much more pronounced than in the models of Sirgue *et al.* (2010) and Prieux *et al.* (2011). Therefore, the relevance of this feature needs further validation. This low-velocity layer might however indicate an accumulation of gas below a lithological barrier (Prieux, 2012). The first frequency group inversion leads to a misfit function reduction of 18%, which was achieved after eight iterations (Figure 3.23).

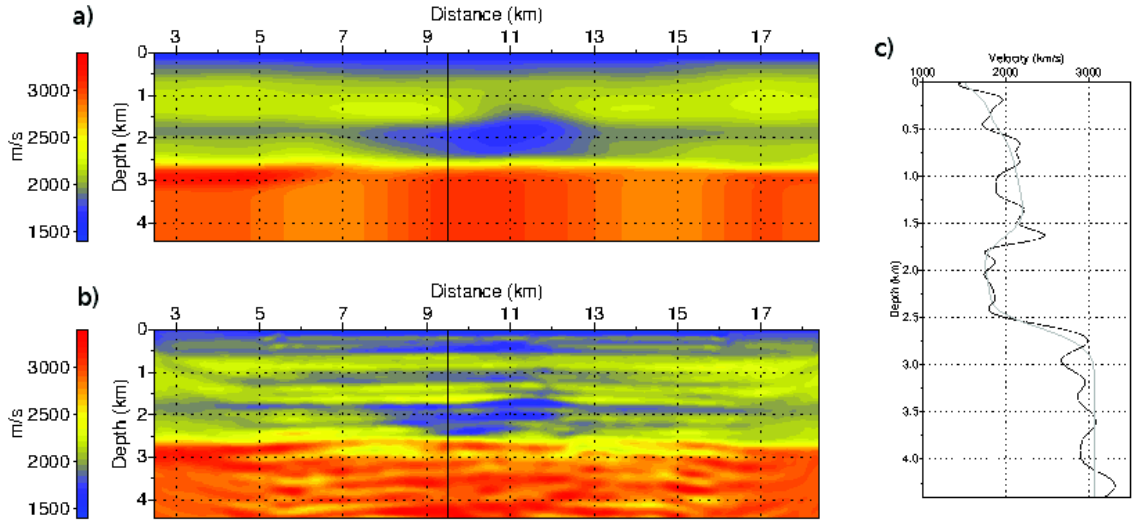

Figure 3.22: Valhall case study - First frequency group inversion: vertical section at x = 6500m extracted from the initial model (a), and from the final FWI model of the first frequency group inversion (b). (c) Vertical profile extracted from the initial (gray line) and the final (black line) FWI models at x = 6500 m and y = 9500 m.

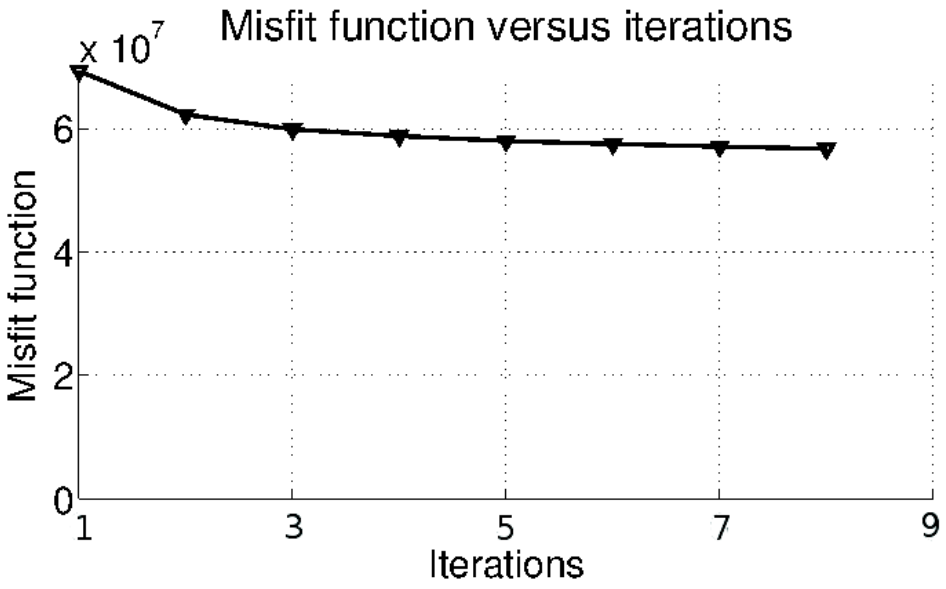

Figure 3.23: Valhall case study - First frequency group inversion. Misfit function versus iterations. A misfit function reduction of 18%, which was achieved after eight iterations. We did not perform too many iterations to avoid matching noise.

We use the final model of the first frequency group ($m8$) as the starting model for the second frequency group. The FWI results of the second frequency group inversion are shown in figures 3.24 and 3.25, with the same presentation than for the first frequency group inversion. Finer structures are resolved in the channels as the inversion progresses towards higher frequencies and the footprint of the acquisition is reduced (note how the polarity of the velocity perturbations associated with the acquisition footprint change between the first and second frequency group inversion). The resolution of the imaging of the gas cloud and of the related gas-filled fractures was also improved after the second frequency group inversion.

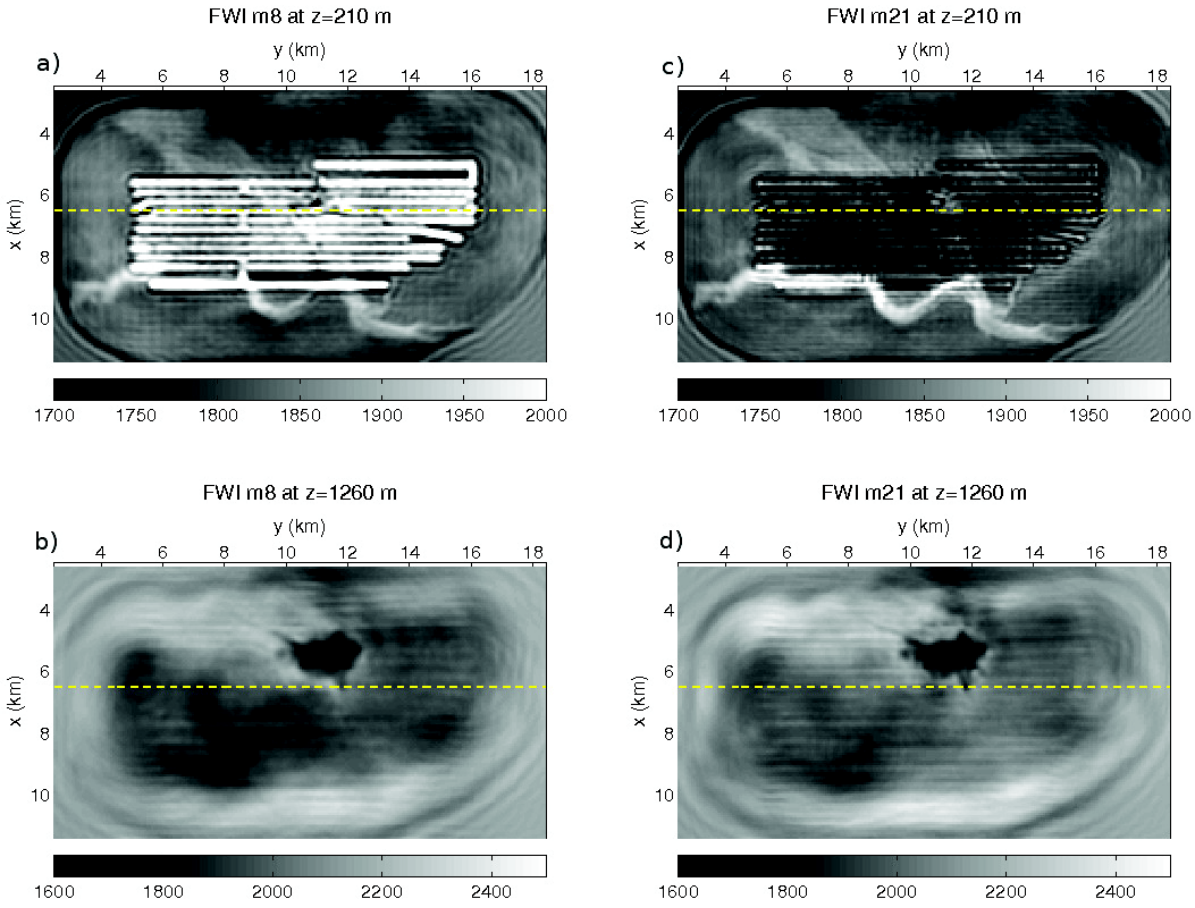

Figure 3.24: Valhall case study - Second frequency group inversion: horizontal slice at a depth of 150 m extracted from the initial model (a), and from the final FWI model of the second frequency group inversion. (c) Horizontal slice at a depth of 1260 m extracted from the initial model (b) and for the final FWI model of the second frequency group inversion (d). The yellow line denotes the position of an available well log.

The vertical section of the FWI model is shown in figure 3.25. Compared to the FWI model obtained after the inversion of the first frequency group, a feature starts appearing at y = 12 km above the gas layers. This feature, which is more visible in the FWI model obtained after the inversion of the third frequency group (Figure 3.28) is interpreted as a gas-filled fracture, which was already identified on the horizontal section of the FWI model (Figure 3.24).The second frequency group leads to a misfit function reduction of 60%, which was reached after 13 iterations. This misfit function reduction is significantly higher than the one obtained for the first frequency group (18%). This simply highlights the higher signal-to-noise ratio of the frequencies of the second frequency group.

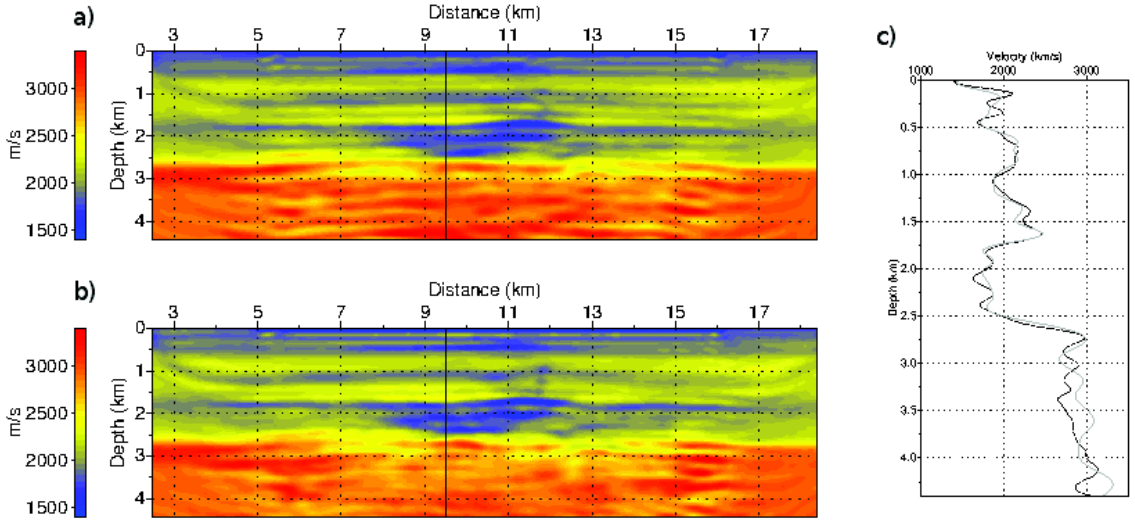

Figure 3.25: Valhall case study - Second frequency group inversion: vertical slices at x = 6500m extracted from the initial model (a), and from the final (b) FWI model of the second frequency group inversion. (c) Vertical profile extracted from the initial (gray line) and the final (black line) FWI models at x = 6500 m and y = 9500 m.

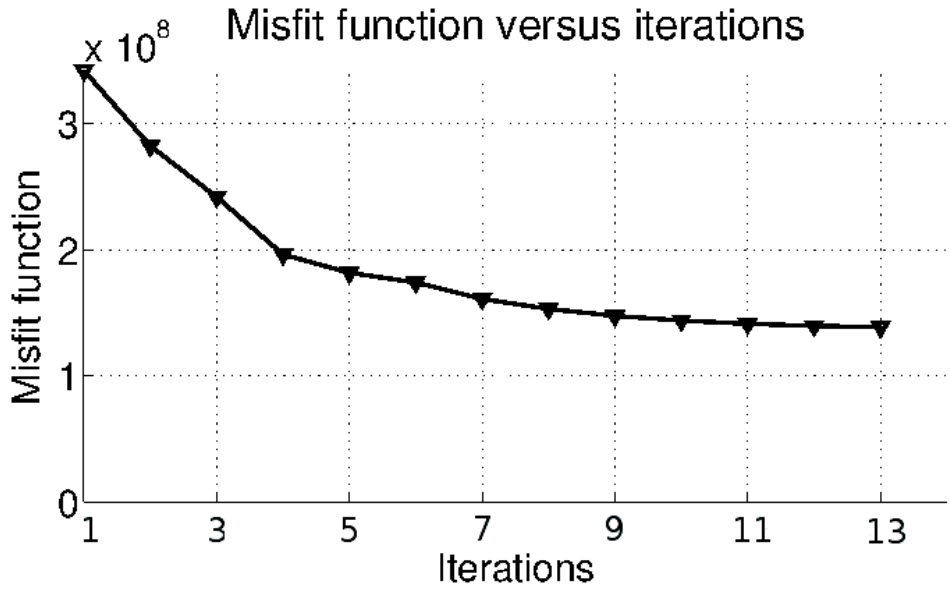

Figure 3.26: Valhall case study - Second frequency group inversion: misfit function versus iteration numbers. A misfit function reduction of 60% was reached after 13 iterations

As for the second group, we use the final model of the second frequency group ($m21$) as the starting model for the third frequency group inversion. The FWI models obtained after the third frequency group inversion are shown in the figures 3.27 and 3.28. The acquisition footprint at 210 m in depth is dramatically reduced, and a very clear picture of the paleo-channels is obtained. The position of the platform is also indicated by a bright spot in the image at x = 6.3 km and y = 11 km. In contrast, the horizontal section of the gas cloud was not significantly improved after the third frequency group inversion. However, as already mentioned, the image of the gas filled fracture at x = 12 km is quite clear in the FWI model of the third frequency group inversion (Figure 3.28).

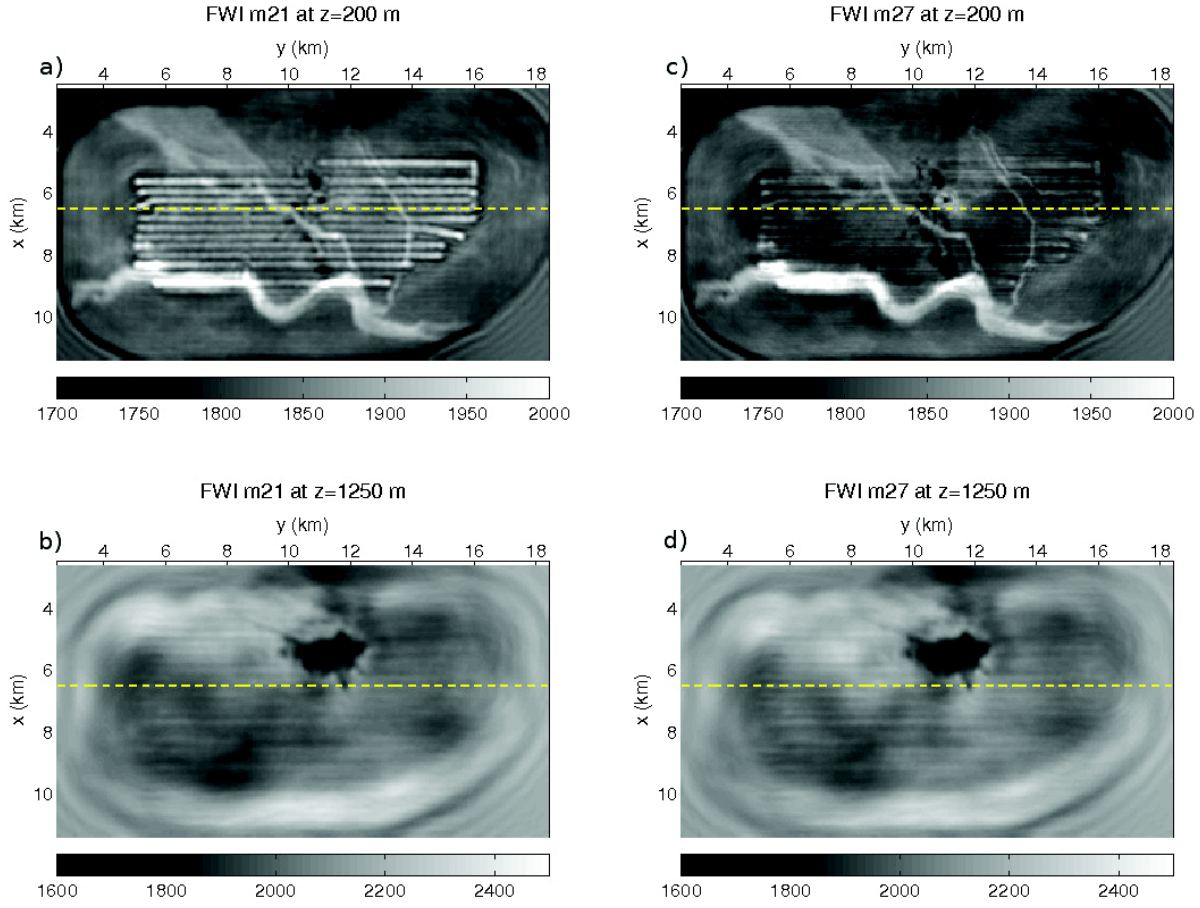

Figure 3.27: Valhall case study - Third frequency group inversion: Horizontal slice at a depth of 150 m extracted from the starting model ($m21$) (a), and from the final FWI model of the third frequency group inversion ($m27$) (c). (b, d) Same as (a, c) for the horizontal slice at 1260 m in depth.

The third frequency group inversion leads to a misfit function reduction of 52%, which was achieved after six iterations (figure 3.29). A comparison between a sonic log located at (x = 6.8 km, y = 9.5 km) and the corresponding vertical profiles extracted from the final FWI model ($m27$) and the initial model($m0$) is shown in the figure 3.30a. The NMO velocities of the initial model are significantly higher than the vertical velocities of the sonic log, which results from the significant anisotropy at Valhall. We show a reasonable agreement between the FWI profile and the sonic log down to 1 km in depth. However, the FWI velocities are significantly higher than the vertical velocities of the sonic log between 1.4 km and 1.8 km in depth. These velocities are close to horizontal velocities, and were built by FWI to match the first arrivals at long offsets (Prieux *et al.*, 2011). At a depth of around 1.2 km, a low velocity zone in the final FWI model has been reconstructed. Although a short-scale low-velocity anomaly is shown in the well log at 1.25 km in depth, the low velocity zone reconstructed by FWI needs further validation, because it is not visible in the 3D FWI model of Sirgue *et al.* (2010) (Figure 3.30).

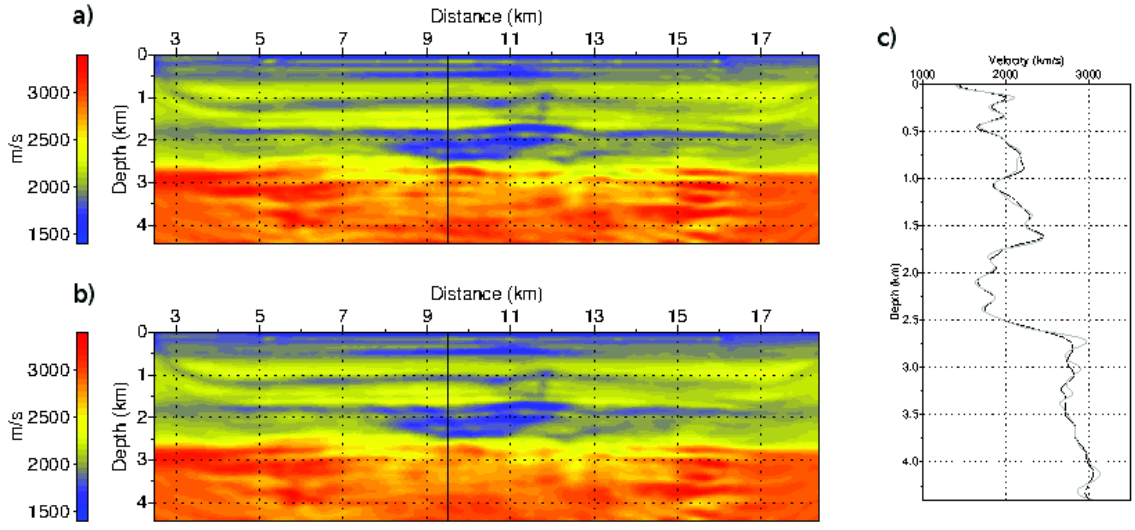

Figure 3.28: Valhall case study - Third frequency group inversion: Vertical slices at x = 6500m extracted from the starting model ($m8$) (a), and from the FWI final of the third frequency group inversion (b). (c) Vertical profile extracted from the initial (gray line) and the final (black line) FWI models at x = 6500 m and y = 9500 m.

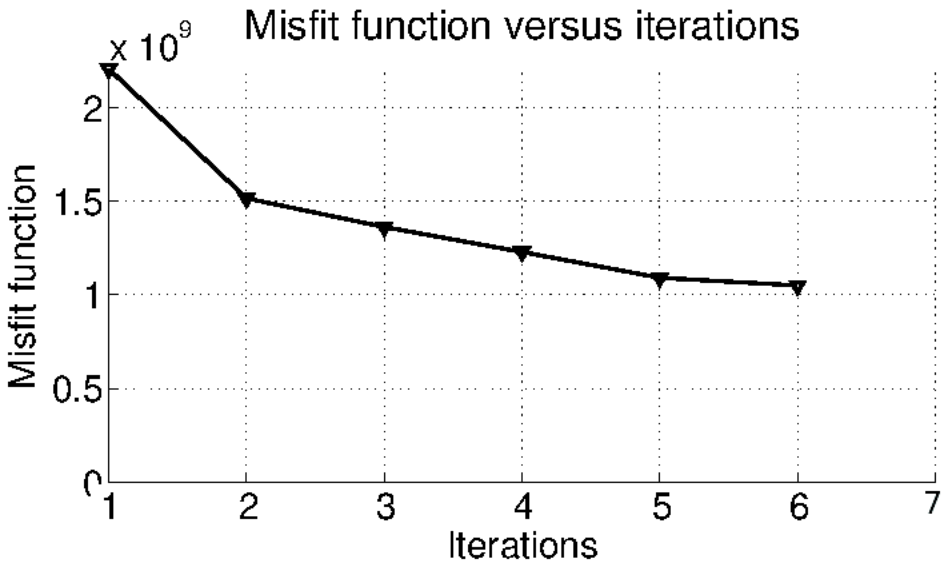

Figure 3.29: Valhall case study - Third frequency group inversion: the misfit function versus iteration number. A misfit function reduction of 52% was achieved after six iterations.

3.2.8 Model appraisals

Model appraisal is a key issue in FWI as uncertainty analysis is quite challenging to perform in a Bayesian framework (Gouveia et Scales, 1998). In this work, the FWI models are assessed based on the local match between a sonic log and the FWI models, the flatness of the common image gathers (CIGs) computed by two-dimensional (2D) reverse time migration (RTM), synthetic seismogram modeling, and by the repeatability of source wavelets, which are estimated for each receiver gather.

3.2.8.1 Comparison with sonic logs

Figure 3.30 shows the comparison between a sonic log and the corresponding vertical profile extracted from the initial model and the final FWI model. Our FWI model shares similar features with the FWI model of Sirgue *et al.* (2010). A low-velocity model perturbation at 0.5 km in depth followed by a progressive increase of velocity with depth down to 0.8 km is a common feature of the sonic log and the FWI models. High velocities, close to horizontal velocities, are reconstructed between around 1.4 km and 2.2 km in depth in our FWI model and the model of Sirgue *et al.* (2010). At these depths, anisotropy is significant in Valhall (Etienne *et al.*, 2012). Prieux *et al.* (2009) proposed that these high velocities were built by FWI to match the first arrivals at long offsets, which mainly propagate in the upper structure, above the gas layers. The high-velocity contrast at the reservoir level at 2.5 km in depth was better delineated by FWI compared to the one shown in the initial model.

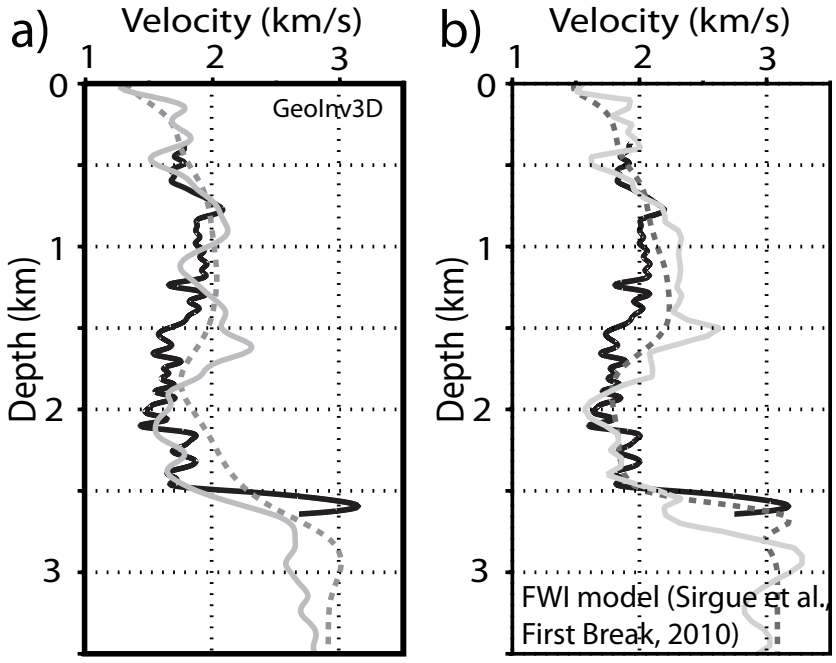

Figure 3.30: Valhall case study - Comparison between sonic logs and FWI profiles: (a) Comparison between the sonic log (black) and the corresponding profiles extracted from the initial model (dashed line) and the final FWI model (solid gray line). (b) Same as (a) but the FWI profile is extracted from the FWI model of Sirgue *et al.* (2010).

3.2.8.2 Synthetic seismogram modeling

We compute synthetic seismograms in the final FWI model of each frequency group inversion. To compute seismograms and compare them with the real data, we need a source wavelet, which is estimated by linear inversion of the full data set using the approach of Pratt (1999). I shall discuss the details of this source estimation in the following section.

We compute time-domain seismograms in the final FWI models of the first, second, and third frequency group inversions (Figures 3.31, 3.33, and 3.34 for a receiver of the cable 21 (the same cable as the one processed by 2D FWI in Prieux *et al.* (2011)). We superimpose on the recorded and modeled receiver gathers the first-arrival traveltimes that are computed in the 3D FWI model with the eikonal solver of Podvin et Lecomte (1991) (Figures 3.31, 3.33, and 3.34, green line).

We first show that the delay between recorded first-arrival traveltimes and modeled first-arrival traveltimes computed in the initial model (Figure 3.13) were removed after the first frequency group inversion (Figure 3.31). However, the superimposition of modeled first-arrival traveltimes with modeled seismograms computed in the FWI model show that the amplitudes of the first-arrival are so weak beyond 5 km in offset (Figure 3.32) than these arrivals are not visible in the seismograms. Only a more aggressive clip of the amplitudes would allow to show this first arrival. This weak first arrival results from the low-velocity layer at around 1.2 km in depth, which was reconstructed in the FWI model (Figures 3.22, 3.25, and 3.28). First-arrival ray tracing performed in a 2D section of the FWI model shows that the rays graze on top of the low-velocity layer, and mimic an evanescent interface wave on top of the low velocity layer (Figure 3.36). The evanescent nature of the first arrival beyond 5 km of offset explains its weak amplitudes in the modeled seismograms. As these weak amplitudes are not shown in the recorded data, we conclude that the negative velocity contrast at 1.2 km in depth in the FWI model was overestimated, a possible footprint of cycle skipping. However, this velocity contrast might be real, although its amplitude was badly estimated. This is supported by the vertical gas-filled fracture at 11 km of distance in figures 3.20 and 3.28, which seems to hallmark this low velocity layer. The low-velocity layer shown in figure 3.28 was interpreted as a gas accumulation between lithological barriers by Prieux (2012) on the basis of elastic FWI results.

Although the waveform of the first arrivals is not fitted well beyond 5 km of offset, we show a good agreement of the phase of the short-spread reflections from the top and bottom of the gas layers (Figures 3.31, 3.33, 3.34, and 3.35). The amplitude of these reflections increases as the inversion progresses towards high frequencies. The poor match of the early arrivals and the good match of the short-spread reflections suggest that FWI is mainly driven by these later arrivals. This may explain why the inversion successfully image reflectors on top and bottom of the gas layers: the reflection response is quite clear in the data (Figure 3.34 bottom).

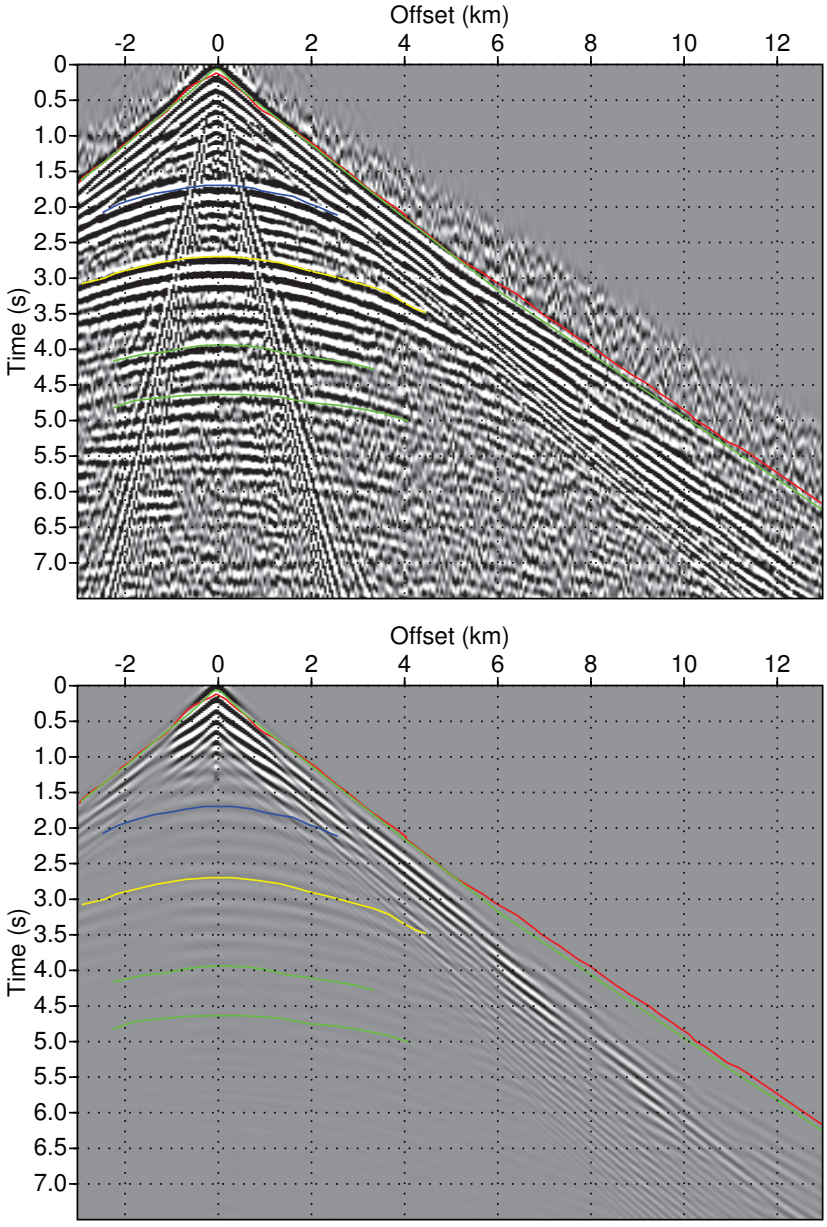

Figure 3.31: Valhall case study - First frequency group inversion: Recorded and modeled receiver gather for shots located along cable 21. Recorded (red line) and modeled (green line) first-arrival traveltimes are superimposed. Picks of reflection traveltimes are also superimposed to facilitate the comparison between the phase of the recorded and modeled seismograms. Note how the delays between the recorded and modeled first-arrival traveltimes beyond 5 km in offset were reduced, compared to modeled seismograms computed in the initial model (Figure 3.13). However, the first-arrivals beyond 5 km in offset have very small amplitudes, and are barely visible in the modeled seismograms. This results because the first arrival is an evanescent interface wave, which propagates on top of the low velocity layer at 1.2 km in depth (Figure 3.36).

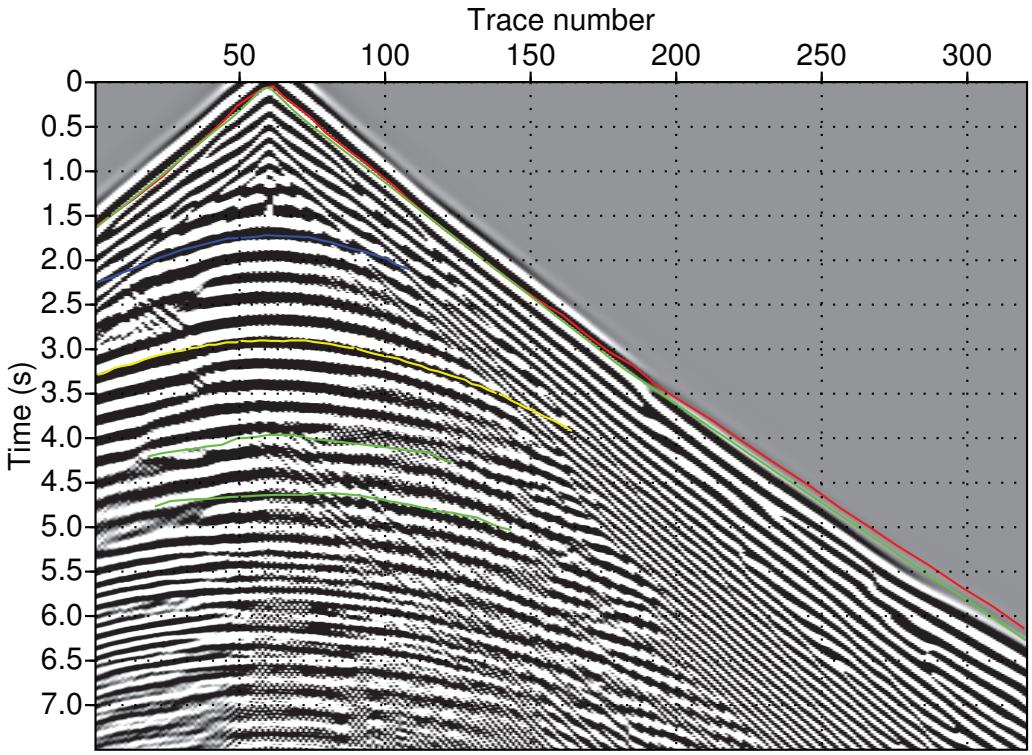

Figure 3.32: Valhall case study - First frequency group inversion: Modeled receiver gather for shots located along cable 21 with the illustration of the weak amplitudes of the first-arrival. Recorded (red line) and modeled (green line) first-arrival traveltimes are superimposed.

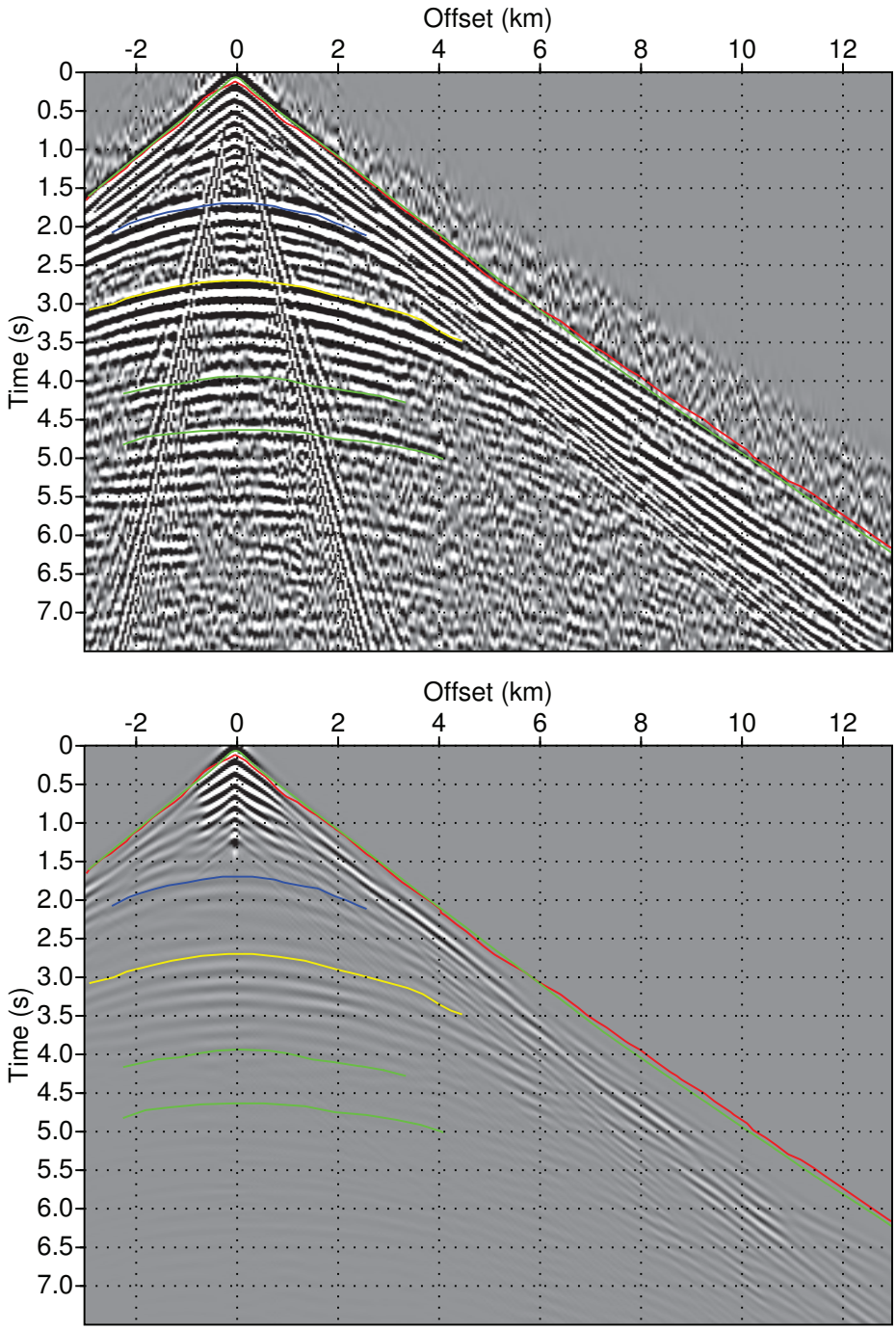

Figure 3.33: Valhall case study - Second frequency group inversion: Same as figure 3.31 but the modeled seismograms are computed in the FWI model obtained close of the second frequency group inversion.

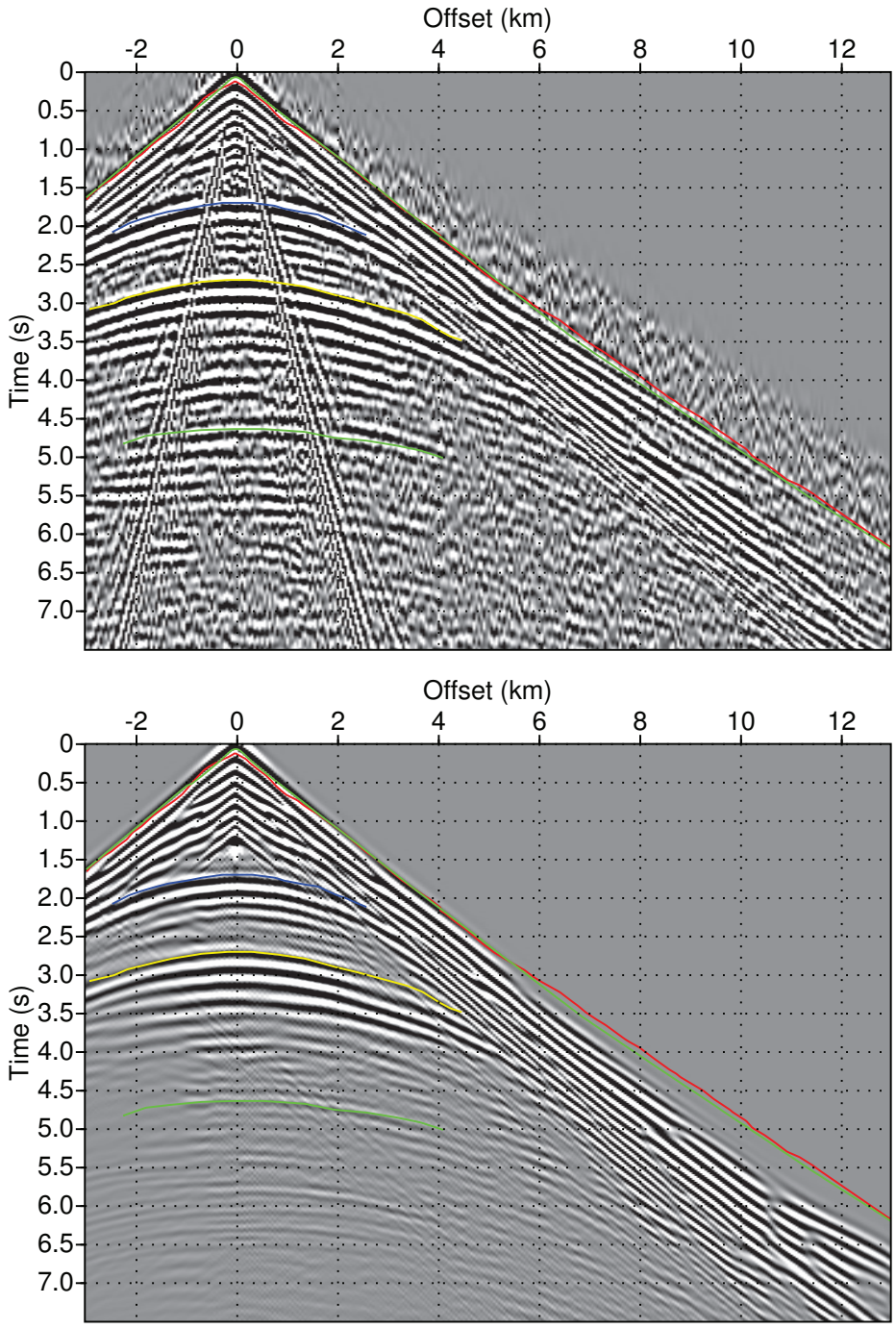

Figure 3.34: Valhall case study - Third frequency group inversion: Same as figure 3.31 but the modeled seismograms are computed in the FWI model obtained close of the third frequency group inversion. Note how the amplitudes of the short-spread reflections were increased compared to the seismograms computed in the FWI models of the first two frequency group inversions.

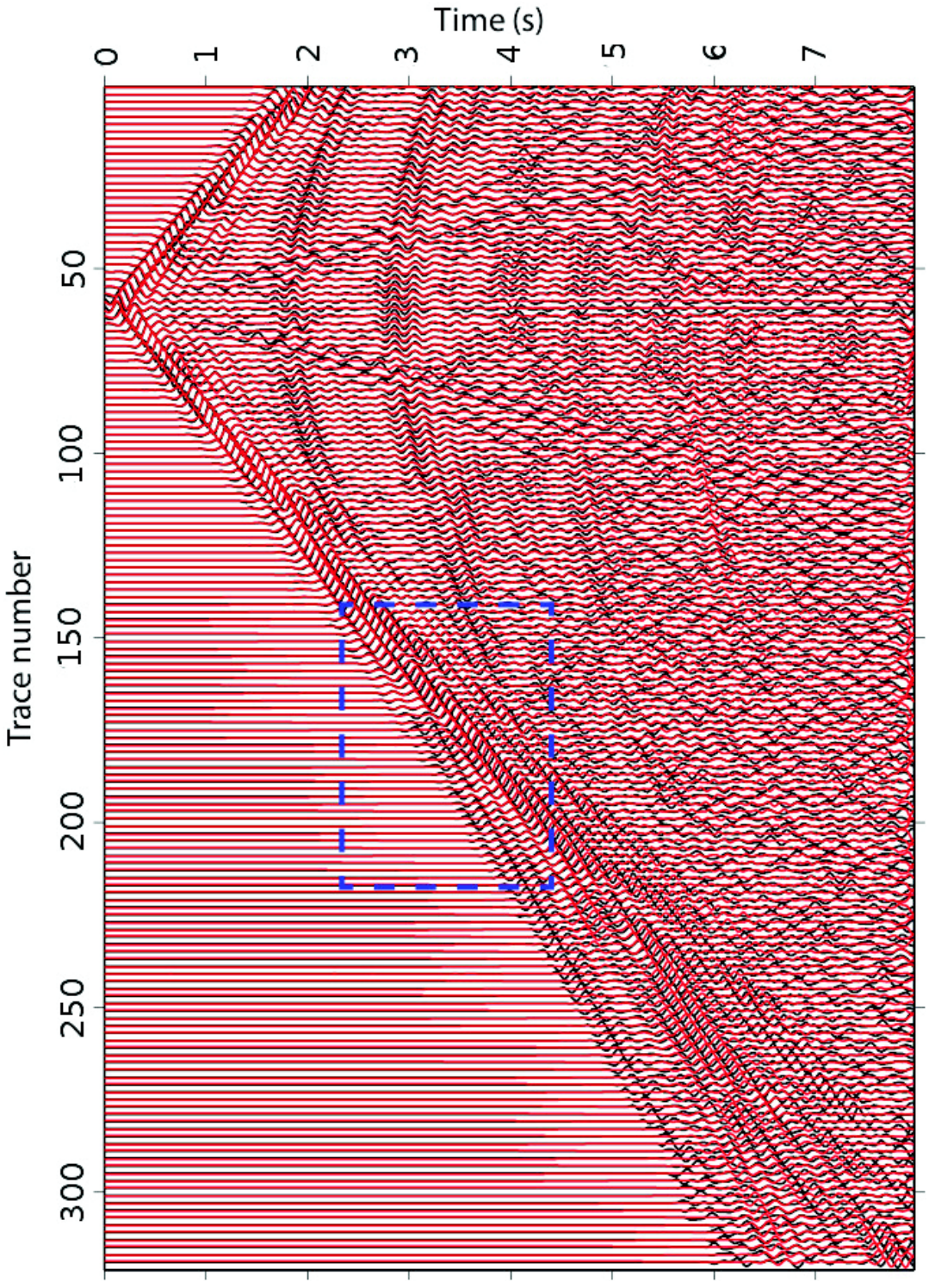

Figure 3.35: Valhall case study - Third frequency group inversion: Direct comparison between recorded and modeled seismograms shown in figure 3.34. An automatic gain control with a window length of 2 s was applied to the recorded and modeled seismograms, because we want to focus on the phase match. The dash box delineates the range of offsets, beyond which we see a phase shift between the recorded and modeled first arrivals.

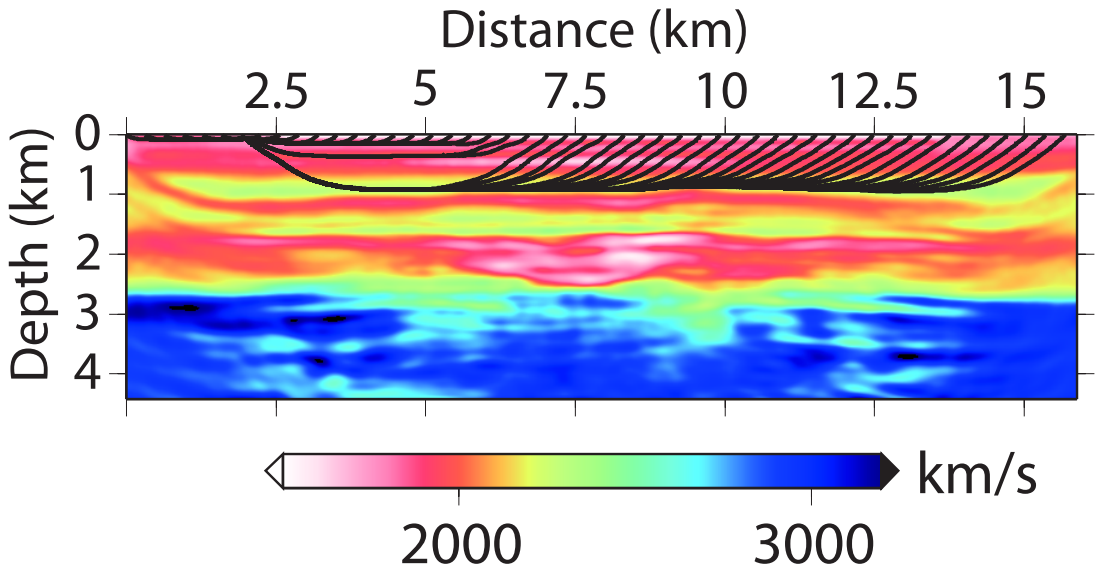

Figure 3.36: Valhall case study - First-arrival ray tracing in the final FWI model. Note how the rays graze on top of the low velocity layer. They represent an evanescent interface wave.

3.2.8.3 Data fit in the frequency domain

For completeness, we also compare recorded and modeled data in the frequency domain. The modeled data computed in the initial model show significant amplitude and phase mismatches with the recorded data for the 4.06 Hz frequency, and the residuals were efficiently reduced after the first frequency group inversion (Figures 3.37 and 3.38). Indeed, the match in the frequency domain of monochromatic data do not guarantee that the inversion converges towards the true model, because of possible cycle skipping artifacts. The same frequency-domain simulation are computed in the initial and final FWI models of the second and third frequency group inversions (Figures 3.39, 3.40, 3.41 and 3.42). The modeled frequencies are 5.04 Hz and 7.01 Hz for the second and third frequency group inversions, respectively. In both case, the misfit was efficiently reduced after FWI.

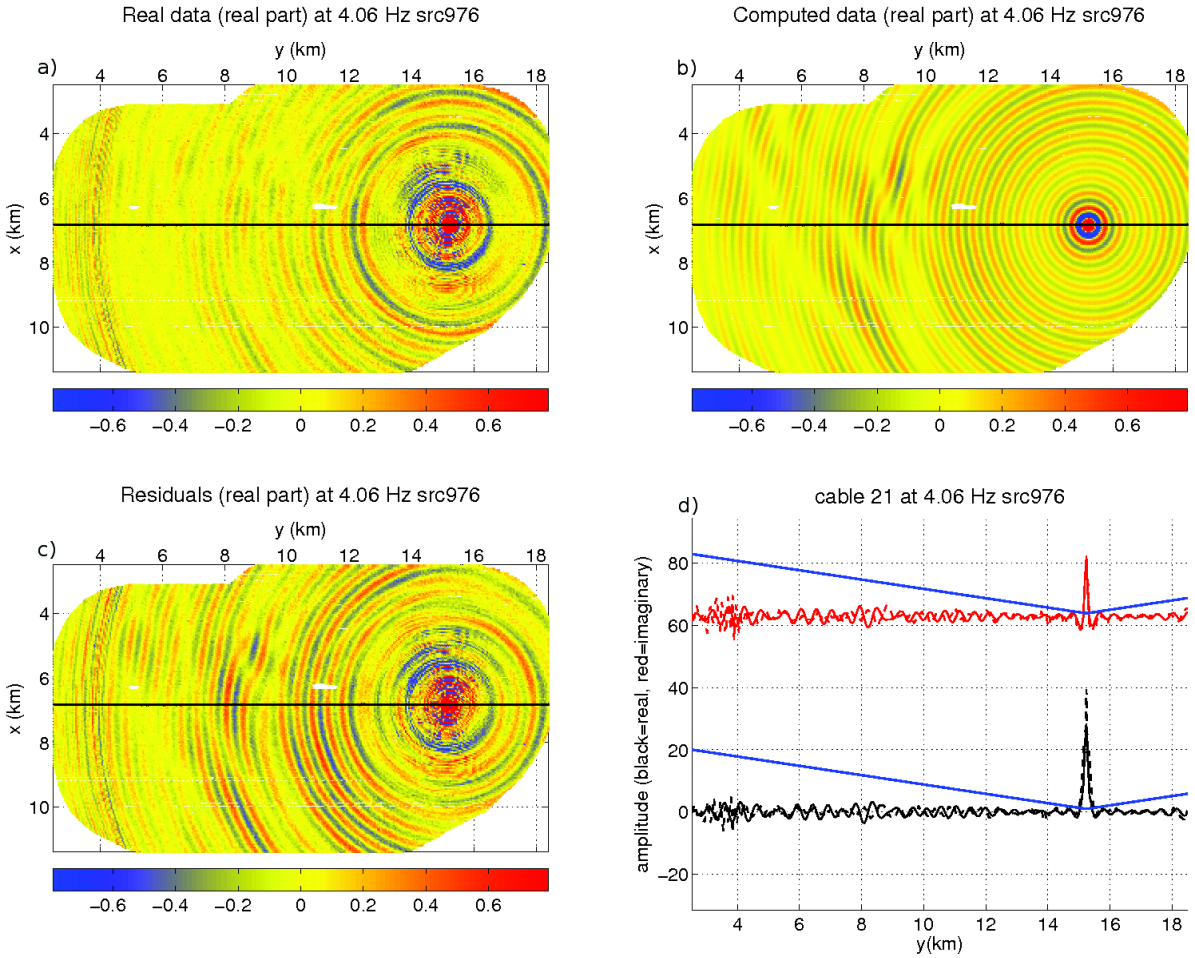

Figure 3.37: Valhall case study - Modeling in the initial model of the first frequency group inversion: (a) recorded monochromatic data (real part) at 4.06 Hz. Source is located at around x=6.8 km and y=15 km; (b) synthetic data; (c) difference between (a) and (b); (d) horizontal profile extracted from the recorded (solid line) and the modeled (dash line) wavefields. Position of the profile is denoted by the black line in (a). The black and red lines correspond to the real and imaginary part of the complex- valued wavefields, respectively. A gain with offset (blue curve) is applied to the amplitudes to assess the match at long offsets.

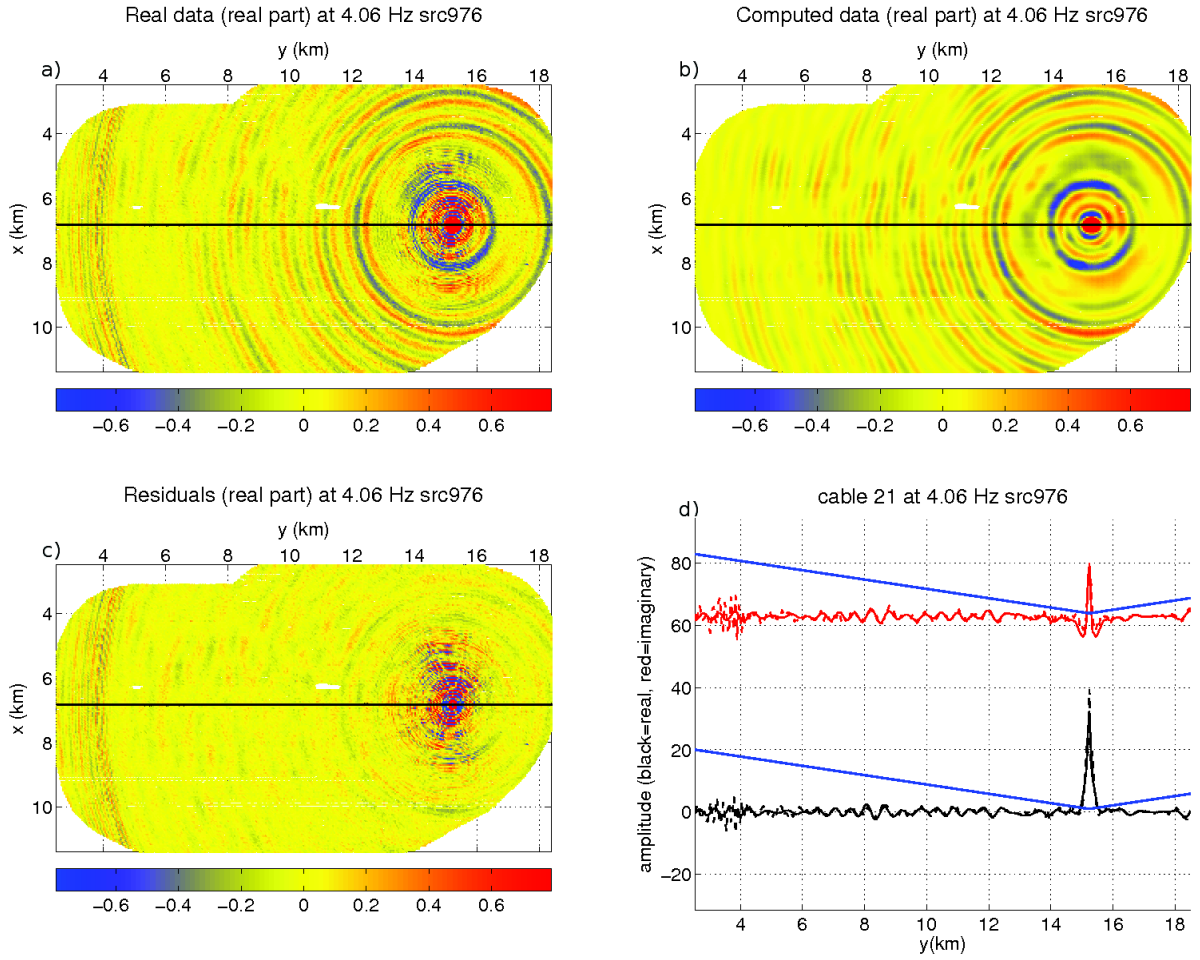

Figure 3.38: Valhall case study - Modeling in the final FWI model of the first frequency group inversion: (a) recorded monochromatic data (real part) at 4.06 Hz. Source is located at around x=6.8 km and y=15 km; (b) synthetic data; (c) difference between (a) and (b); (d) horizontal profile extracted from the recorded (solid line) and the modeled (dash line) wavefields. Position of the profile is denoted by the black line in (a). The black and red lines correspond to the real and imaginary part of the complex- valued wavefields, respectively. A gain with offset (blue curve) is applied to the amplitudes to assess the match at long offsets.

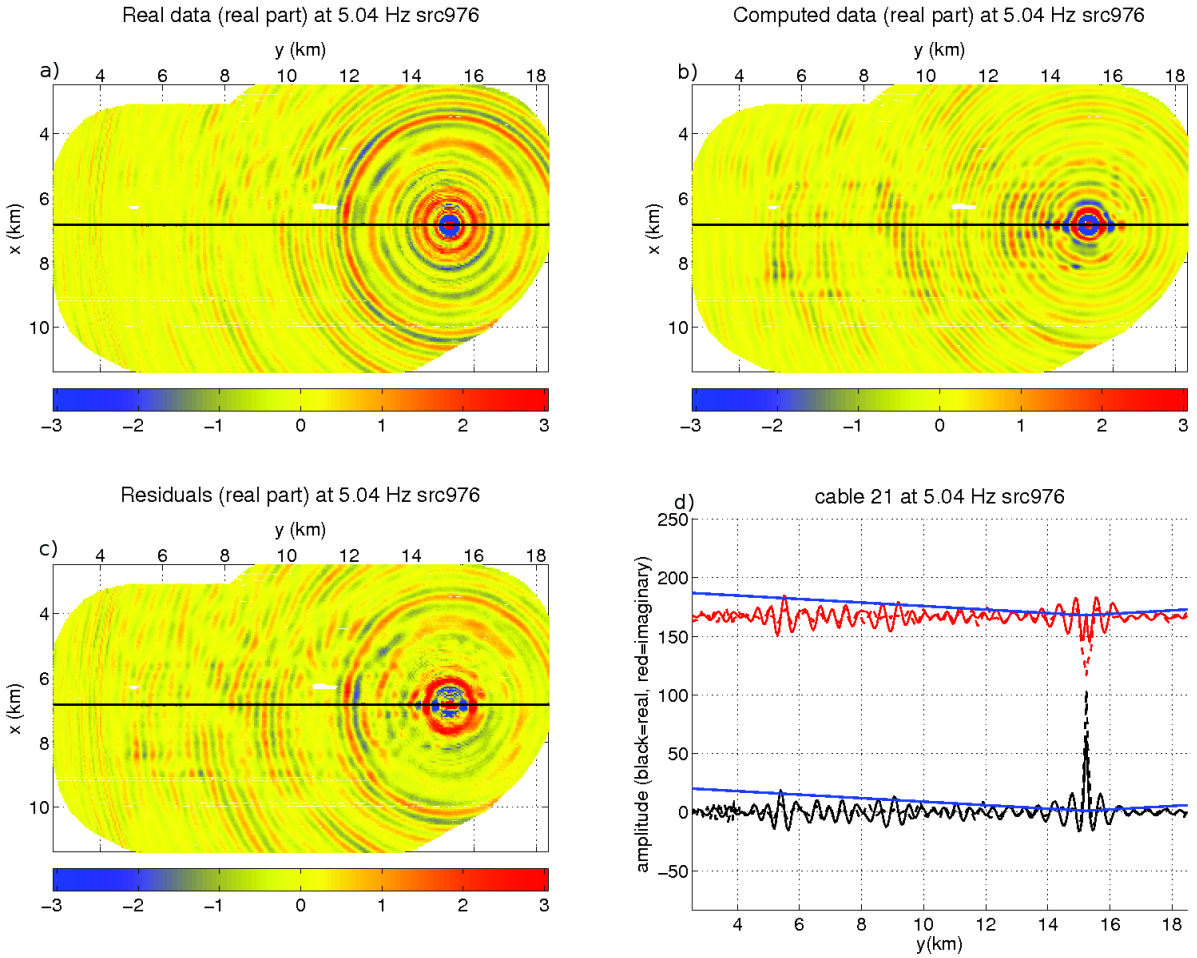

Figure 3.39: Valhall case study - Modeling in the initial model of the second frequency group inversion: (a) recorded monochromatic data (real part) at 5.04 Hz. Source is located at around x=6.8 km and y=15 km; (b) synthetic data; (c) difference between (a) and (b); (d) horizontal profile extracted from the recorded (solid line) and the modeled (dash line) wavefields. Position of the profile is denoted by the black line in (a). The black and red lines correspond to the real and imaginary part of the complex-valued wavefields, respectively. A gain with offset (blue curve) is applied to the amplitudes to assess the match at long offsets.

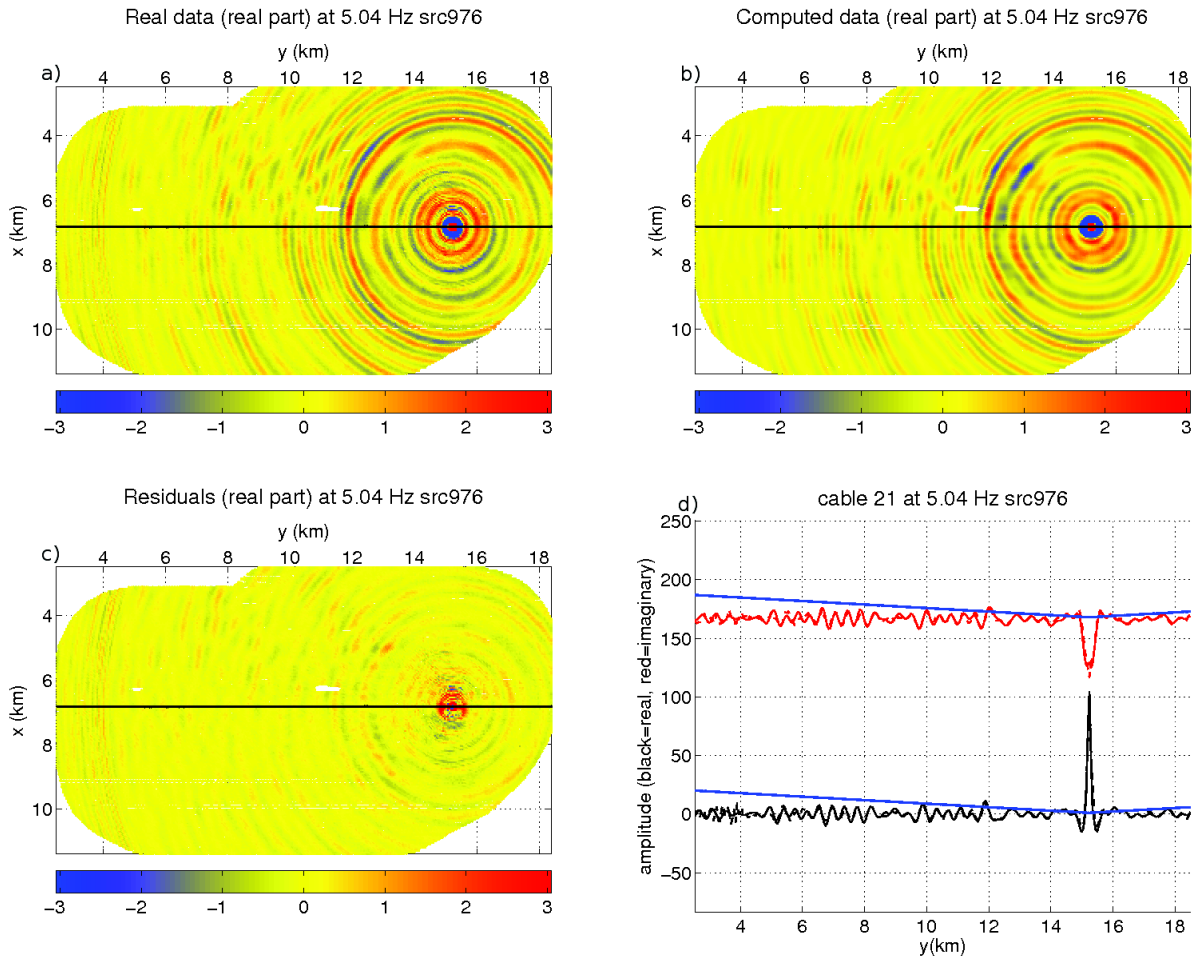

Figure 3.40: Valhall case study - Modeling in the final FWI model of the second frequency group inversion: (a) recorded monochromatic data (real part) at 5.04 Hz. Source is located at around x=6.8 km and y=15 km; (b) synthetic data; (c) difference between (a) and (b); (d) horizontal profile extracted from the recorded (solid line) and the modeled (dash line) wavefields. Position of the profile is denoted by the black line in (a). The black and red lines correspond to the real and imaginary part of the complex- valued wavefields, respectively. A gain with offset (blue curve) is applied to the amplitudes to assess the match at long offsets.

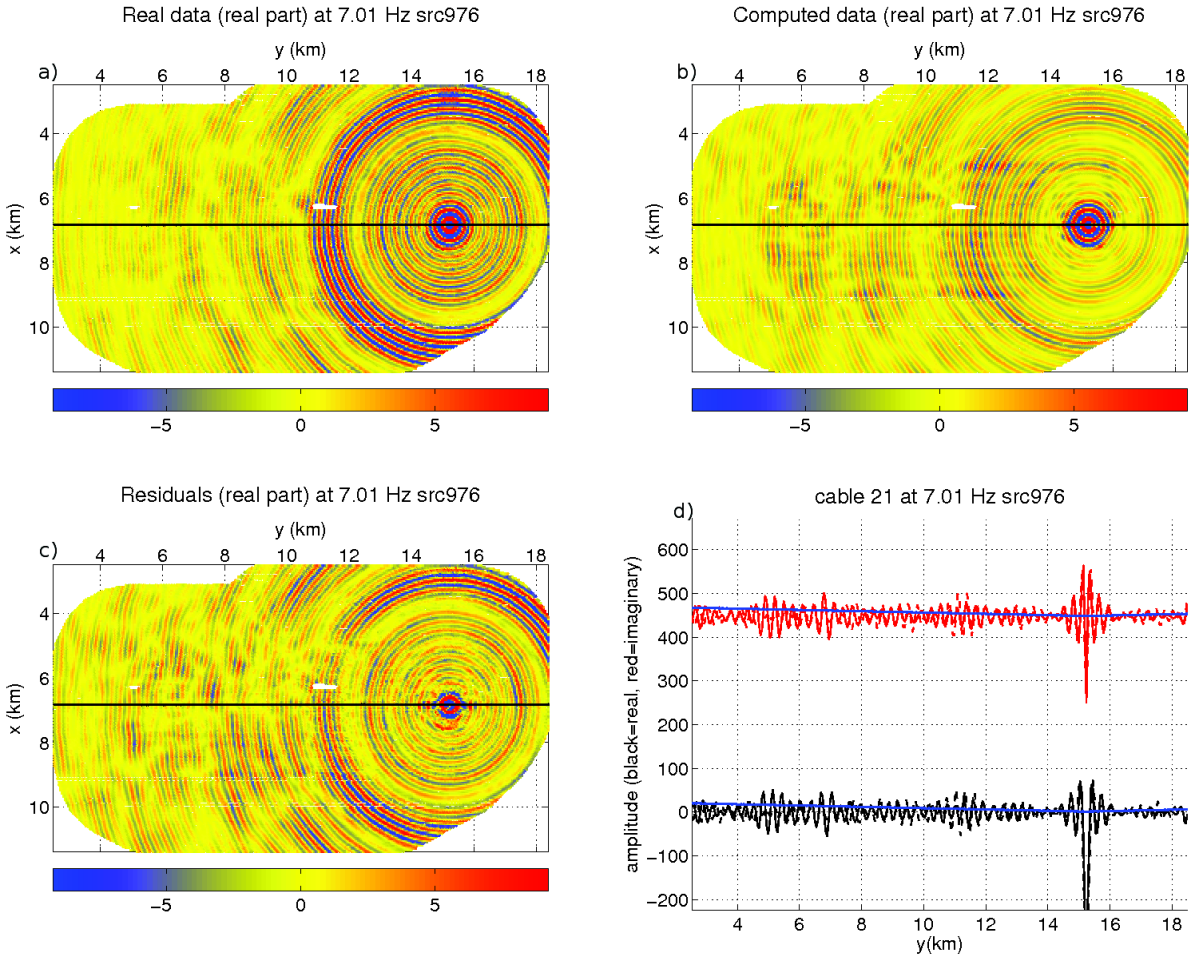

Figure 3.41: Valhall case study - Modeling in the initial model of the third frequency group inversion: (a) recorded monochromatic data (real part) at 7.01 Hz. Source is located at around x=6.8 km and y=15 km; (b) synthetic data (c) difference between (a) and (b); (d) horizontal profile extracted from the recorded (solid line) and the modeled (dash line) wavefields. Position of the profile is denoted by the black line in (a). The black and red lines correspond to the real and imaginary part of the complex- valued wavefields, respectively. A gain with offset (blue curve) is applied to the amplitudes to assess the match at long offsets.

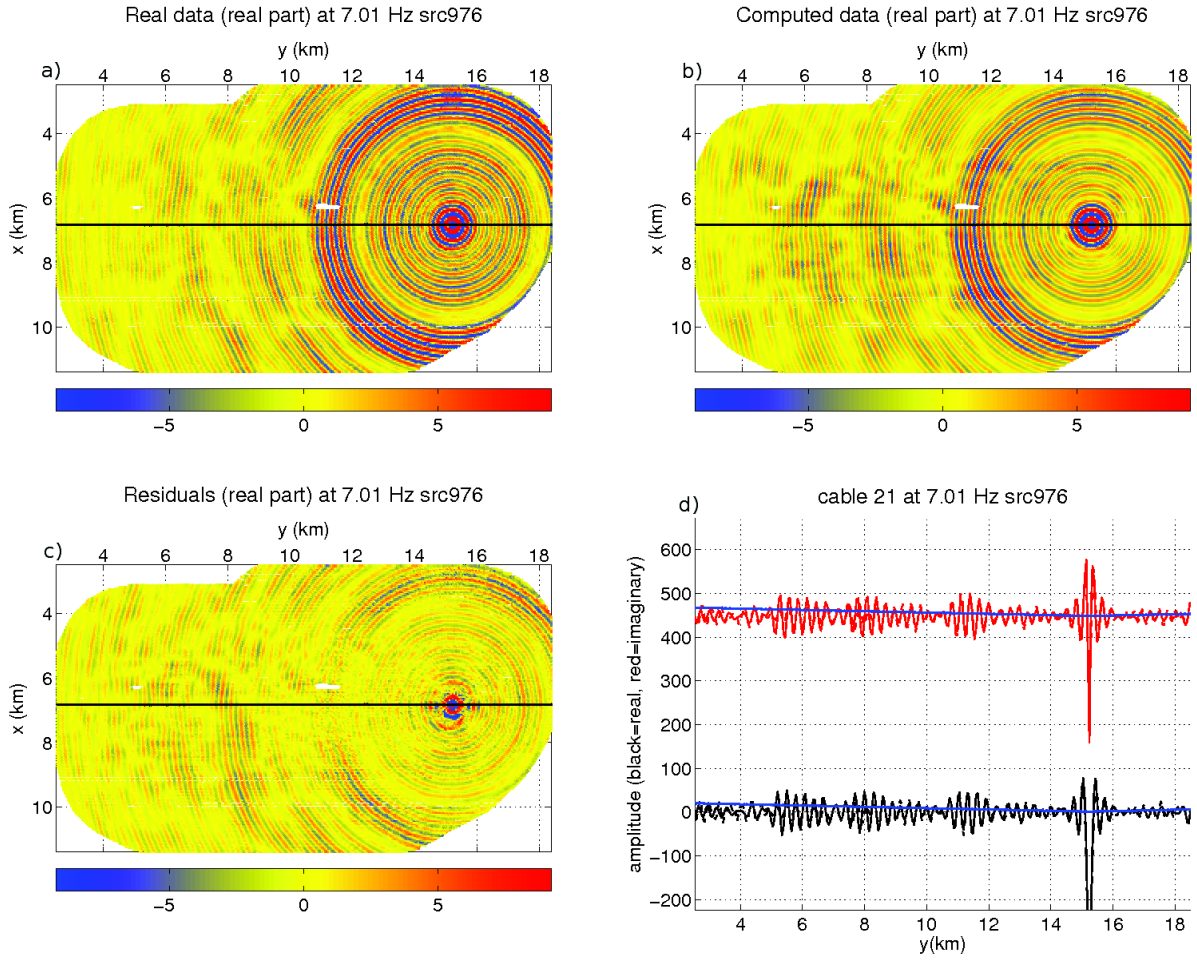

Figure 3.42: Valhall case study - Modeling in the final model of the third frequency group inversion. (a) recorded monochromatic data (real part) at 7.01 Hz. Source is located at around x=6.8 km and y=15 km; (b) synthetic data; (c) difference between (a) and (b); (d) horizontal profile extracted from the recorded (solid line) and the modeled (dash line) wavefields. Position of the profile is denoted by the black line in (a). The black and red lines correspond to the real and imaginary part of the complex- valued wavefields, respectively. A gain with offset (blue curve) is applied to the amplitudes to assess the match at long offsets.

3.2.8.4 Source wavelet estimation

We use the source wavelet estimation as a tool to appraise the relevance of the FWI models (Brenders et Pratt, 2007; Jaiswal *et al.*, 2009; Malinowski *et al.*, 2011; Prieux *et al.*, 2011). The source wavelet is estimated within a frequency band between 0 Hz and 14.1 Hz (consider the data analysis figure 3.10c). As we use a grid interval at 50 m for modeling, the source estimation may be hampered by numerical dispersion, as dispersion is expected for frequencies greater than 7.5 Hz for a minimum velocity of 1500 m/s. We estimate one source wavelet per receiver-gather in the initial model and in the final FWI model of each frequency group inversions (Figure 3.43). As the source estimation is used as a tool to appraise FWI models, all of the shots of the experiment (i.e., all of the offsets) are used to make the source estimation as sensitive as possible to the accuracy of the initial model. 2297 source wavelets are inferred from the initial model ($m0$) and the final model FWI of each frequency group ($m8$, $m21$, $m27$), shown in figure 3.43(a-d, left part) . A mean wavelet is shown in the figure 3.43 (right part). Figure 3.43d shows that the source wavelets inferred from initial model are not optimally focused. As the accuracy of the velocity model improves after FWI of increasing frequencies, wavelets become more and more repetitive, and the amplitude of the mean wavelet increases. Of note, as no offset and time preconditioning is applied to the data, the wavelets are mainly reconstructed from the high-amplitude short-offset arrivals. To increase the sensitivity of the source wavelet estimation to the accuracy of the velocity model, a gain with offset could have been applied to the data to strengthen the contribution of the long offsets.

Time-domain seismograms are computed with a Dirac wavelet in the initial and final FWI models. Although numerical dispersion is visible in the seismograms associated with the Green functions, convolution of these seismograms with the limited-bandwidth source wavelet removes this numerical dispersion. This procedure, which allows to rapidly compute several datasets for different wavelet sources by convolution, is illustrated in figure 3.44. The modeled seismograms after convolution with the source wavelet are compared with the recorded seismograms in figures 3.31, 3.33, and 3.34).

3.2.8.5 Reverse time migration

We compute two-dimensional reverse time migration (RTM) and common-image gathers (CIGs) in the offset-depth domain along cable 21 (Figures 3.45 and 3.46). RTM is performed in two dimensions, because three-dimensional data preprocessed for RTM were not available to us. The reverse time migrated image is computed in the final FWI model, and for a dataset pre-processed for seismic reflection imaging (Prieux *et al.*, 2011). RTM is performed in the frequency domain using the acoustic VTI finite-difference frequency-domain modeling method of (Operto *et al.*, 2009) and the gradient of the FWI program of (Sourbier *et al.*, 2009a,b), where the data residuals are replaced by the data. Each of the common-offset migrated images were computed independently to generate CIGs before stacking. The RTM computed in the two models show structures at similar depths, although the top of the reservoir and the deep reflector between 3.5 km and 4 km in depth are better focused in the RTM image computed in the FWI model of Sirgue *et al.* (2010).

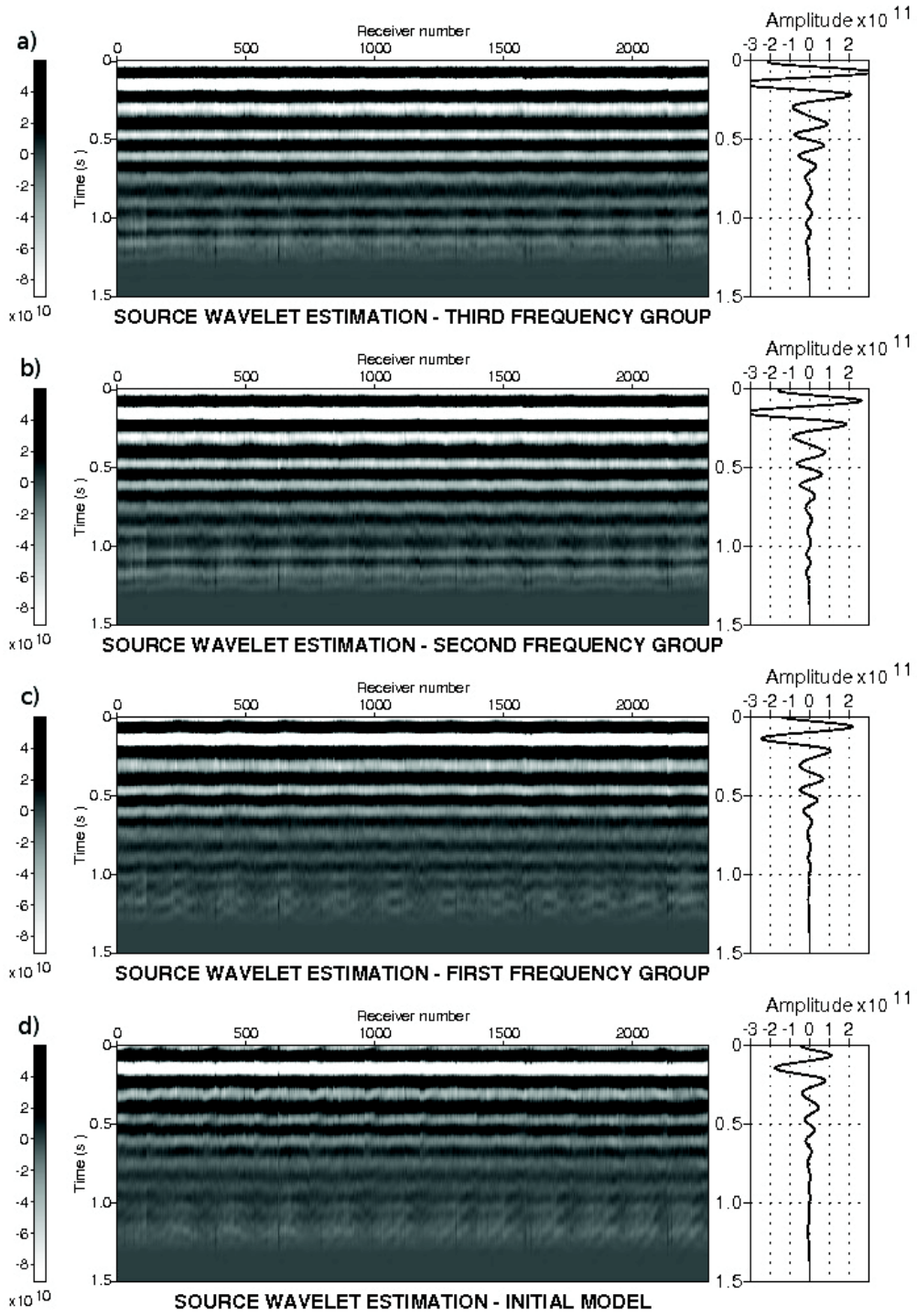

Figure 3.43: Source wavelets per each receiver gather estimated in initial (d) and the final model of first (c), second (b) and third (a) frequency group. On the right, the corresponding mean wavelets.

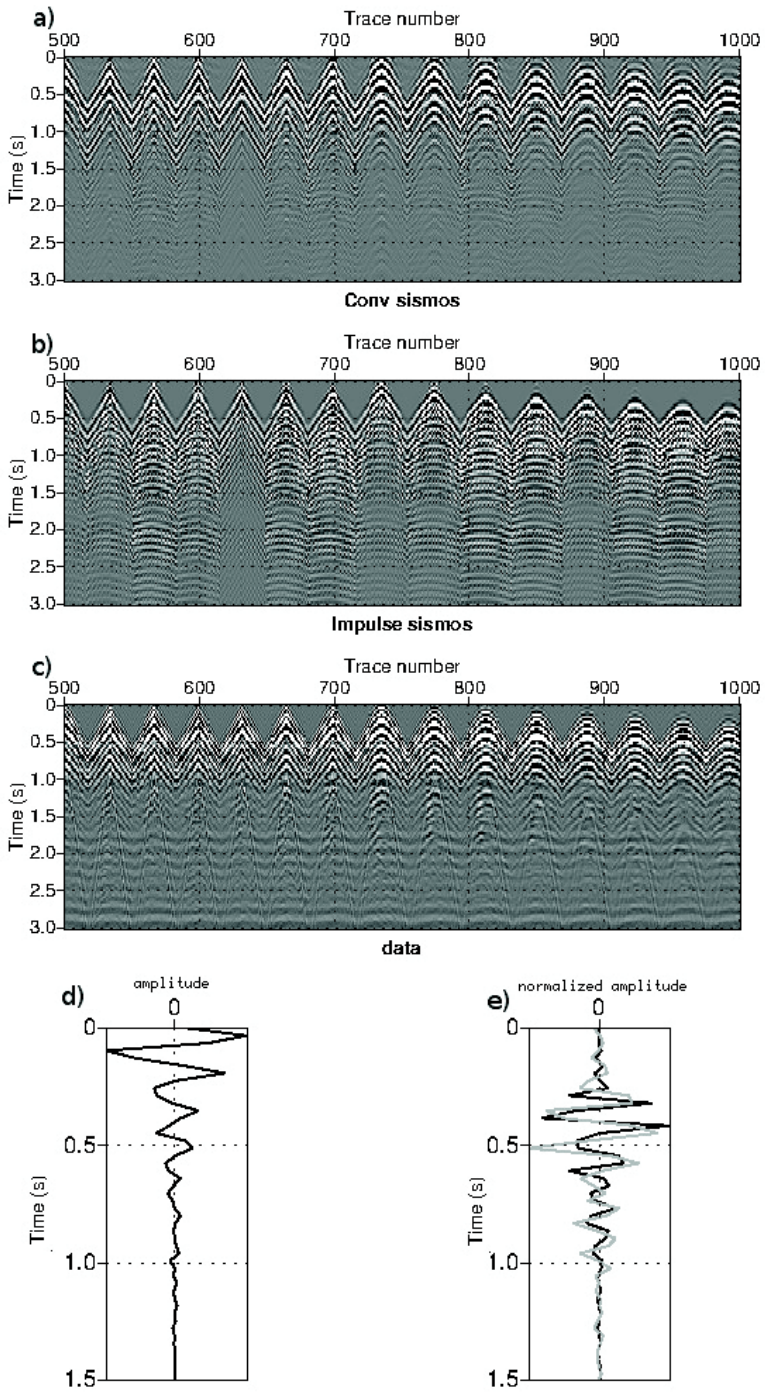

Figure 3.44: Initial model appraisal with source estimation: part seismogram for one receiver gather (cable 21, - receiver $N^0 15$), The recorded data (c), the estimated source wavelet (d), the computed data with the source Dirac (b), the seismogram convoluted with the estimated source wavelet and the log comparison between the two seismograms (e). A good agreement at the shallow and, significant phase mismatches at the depth.

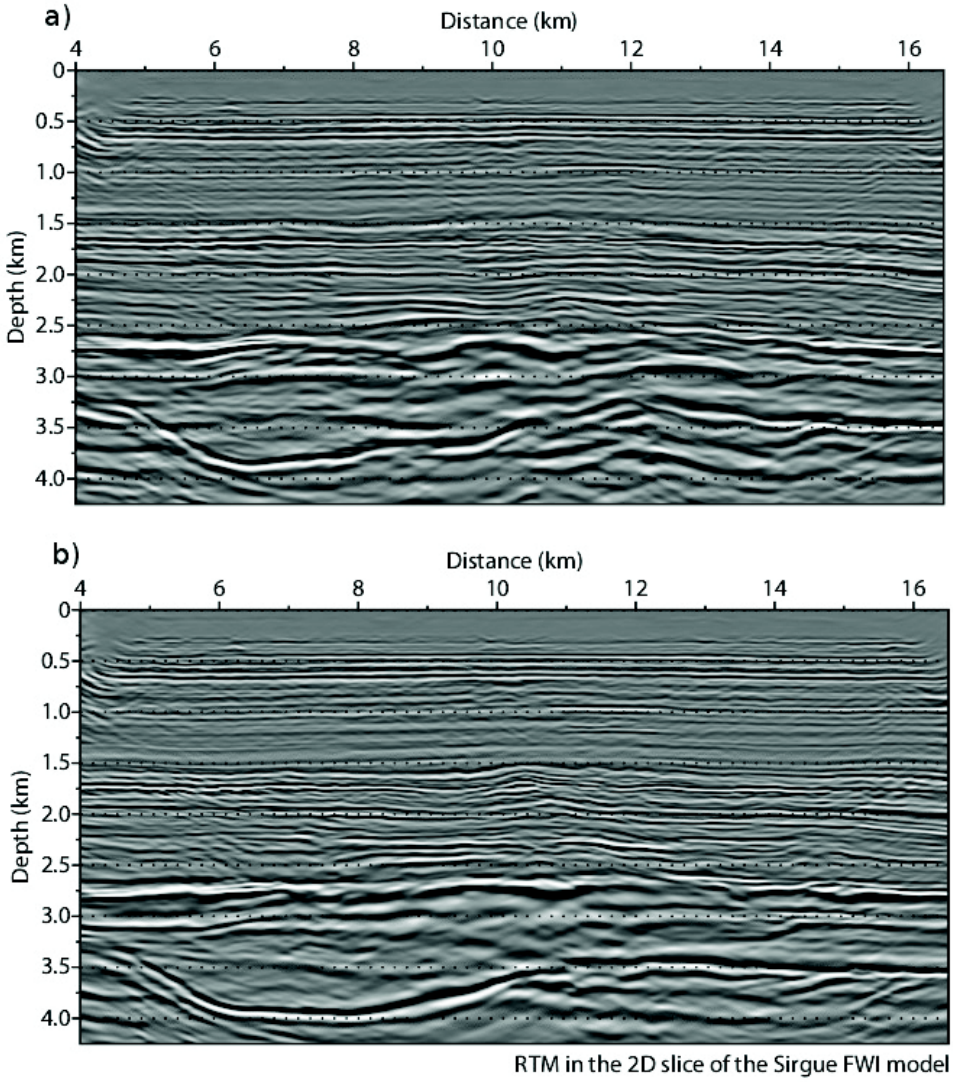

Figure 3.45: RTM image: (a) computed in 2D section of our FWI model (x = 6500 m). (b) the some section computed from Sirgue *et al.* (2010). A quite similar results, (b) is few better in the bottom.

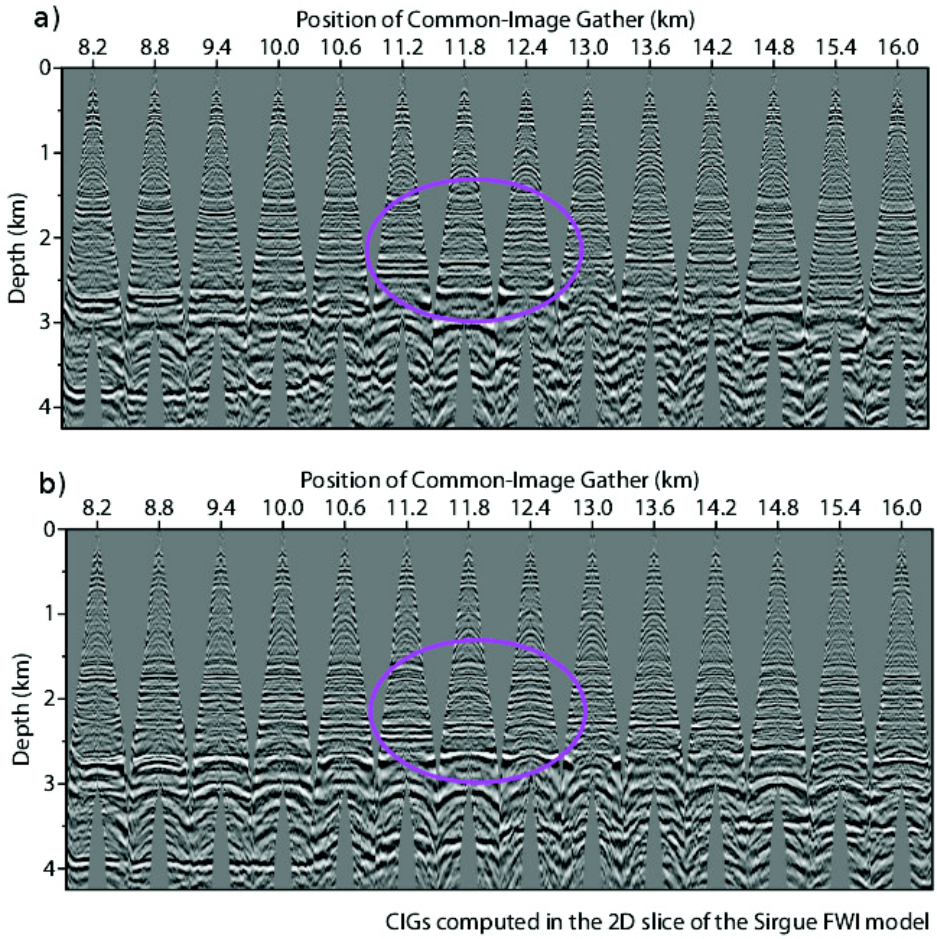

Figure 3.46: CIGs corresponding to RTM images of figure 3.45. The horizontal axis gives the horizontal position of the CIG. The offset range is from -5 km to 5 km. Internal and external mutes ware applied to the CIGs, which are plotted with an automatic gain control. We observe some reflector are more flat (i.e. pink circle).

3.3 Partial conclusion for applications

I presented an application of our 3D FWI code *GeoInv3D* to the EAGE/SEG Overthrust synthetic model and wide-aperture OBC hydrophone data from the Valhall field. I first present the results of the synthetic example to validate the code on a model and data set of representative size. However, as I used the same forward modeling engine for synthetic observed data and for synthetic modeled data, the inversion results are not hampered by usual source of errors such as noise and approximation in the wave physics.

For the real data case study, I showed that the final model of the Valhall application provides a high-resolution seismic velocity model. The initial model used for the inversion was built by anisotropic reflection traveltime tomography before conversion into NMO velocity for isotropic FWI application. This velocity model does not predict sufficiently accurately the first-arrival traveltimes at long offsets to guarantee a reliable inversion of the early arrivals. Although FWI succeeds in reconstruction real features from the reflection wavefields, artifacts might have been introduced in the subsurface model. This study suggests that a careful quality control of the FWI results should be performed when wide-azimuth data are inverted in the isotropic approximation to assess which part of the wavefield has been successfully explained by FWI.

The initial frequency that was used in this study is 3.5 Hz. Lower frequencies would have been useful to reduce the risk of cycle skipping artifacts at long offsets. The importance of low frequencies in FWI of long-offset land data was recently illustrated by Plessix (2012), and this prompts the oil industry to develop new technologies to record them (Soubaras et Whiting, 2011).

Although the Valhall field is known to be anisotropic, we performed FWI in the isotropic approximation. (Prieux *et al.*, 2011) showed the the footprint of anisotropy on isotropic FWI of wide-aperture data. Future work will require to implement anisotropy in the modeling and inversion schemes of *GeoInv3D* code before considering 3D elastic FWI

Chapter 4

Conclusions and perspectives

Contents

The objective of this thesis is the development of a methodology of three-dimensional (3D) acoustic imaging by frequency domain full waveform inversion (FWI) based on time domain modeling with an application to the Valhall ocean-bottom-cable (OBC) data. In order to achieve this goal, I elaborated tools with the help of V. Etienne (Research Engineer, Geoazur) for wave modeling and inversion. I designed and applied our 3D acoustic FWI code to several case studies for the validation of the effectiveness and efficiency of our implementation both for modeling and inversion.

4.1 Forward problem

The first part is devoted to the forward problem: I introduced the wave propagation in the time domain for numerical modeling. I have discussed the pseudo-conservative form of the wave equation, which makes the implementation of adjoint state method straightforward. I chose a conventional ($\mathcal{O}(\Delta x^4, \Delta t^2)$) finite-difference time-domain staggered grid method for its its simplicity and efficiency. The accuracy of the numerical solution requires five grid points per wavelength. This approach is competitive thanks to its high degree of adaptability, the simplicity of implementation, and good scalability on distributed memory architectures.

To deal with the realistic set ups, where sources and receivers need to be accurately implemented on coarse grids, a non-linear interpolation based on sinc function (*Hicks* interpolation) is implemented. This is particularly important when we apply FWI to real data at low frequencies and when the sources and receivers are close to the free surface. I validated the implementation of the sources and receivers against analytical solutions.

The forward modeling is parallelized by domain decomposition on Cartesian grids. One subdomain is assigned to one message-passing-interface (MPI) process. Therefore, the memory

demand and time complexity can be significantly reduced on one MPI process. If multi-source modeling must be performed as in FWI, combining this parallelism with a parallelism over shots makes 3D FWI feasible with present computer resources. These two levels of parallelism can be combined during one single run using two MPI communicators. The first one manages single-source modelings, which are performed in parallel through domain decomposition on a group of processors. The second communicator aim to stack through collective communications the gradients associated with the different sources.

4.2 Inverse problem

In the second part of this thesis, I reviewed 3D acoustic frequency-domain FWI. FWI is formulated as a least-squares minimization problem which tries to minimize the misfit between the recorded and calculated seismic data. The FWI needs an efficient local differential approach to estimate the gradient and the Hessian. The local optimization does not prevent convergence of the misfit function toward local minimum, because of the limited accuracy of the starting model, the lack of low frequency, the noise, and the approximate modeling of the wave-physics complexity. A hierarchical multiscale approach successively inverts data subsets of increasing resolution power to mitigate the non-linearity and ill-posedness of FWI by incorporating shorter wavelengths in the parameter space.

In order to solve the inverse problem, I have linearized the problem by using a quadratic approximation of the misfit function. The optimization algorithm is based on a preconditioned conjugate gradient method. I presented the approach that was designed to compute the gradient of the misfit function with the adjoint-state method. The gradient is preconditioned by the inverse of the pseudo Hessian. The gradient inferred from the adjoint state method was validated against the finite-difference method.

One of the main problems in FWI is the computational burden of multi-source. FWI algorithms must be implemented in parallel to address this problems. I have shown the efficiency of the parallelism, which was implemented in the FWI code.

The FWI code was validated against several synthetic tests, performed in a target of the SEG/EAGE overthrust model. Both marine and land models were considered for this validation.

4.3 The Valhall real data case study

The application of 3D acoustic isotropic FWI to wide-azimuth OBC data from the Valhall oil field is the main goal of this thesis. An application to the full EAGE/SEG Overthrust validation model was achieved before the Valhall real data application, to check the performances of the FWI code on a problem of representative size. Taking into account the limited computer resources, I considered a maximum frequency of 7Hz. The FWI models provide high-resolution images of the subsurface, although higher frequencies would allow to still increase the resolution. This synthetic experiment was performed with the inverse crime, namely, the same modeling engine is used to generate the data and to perform FWI. Therefore, the footprint of many sources of errors were not investigated, and this investigation would deserve further studies.

In the final part of this thesis, I presented the application to the real data set from Valhall. I successively inverted three overlapping frequency groups up to 7 Hz. The initial model was built by vertical transverse isotropic (VTI) reflection traveltime tomography, and was subsequently converted in normal-moveout (NMO) velocity and smoothed for isotropic FWI. Even though the starting model did not contain any hint of channel presence, FWI succeeded in detecting the complex network of channels in the near surface. The potential benefit of long offset data in terms of aperture illumination to reconstruct the deep part of the model below the gas cloud is illustrated. I have successfully imaged several features such as gas cloud with possible gas-filled fractures from the reflection wavefields.

A good agreement between the final FWI model and the sonic log has been observed down to 1 km and a low-velocity layer has been shown at around 1.2 km in depth. We showed that the initial model is not accurate enough to match first-arrival traveltimes at long offset: the time delay reaches a maximum value of 0.24s, which exceeds half of the period (0.14 s) of the starting frequency (3.5 Hz). The inversion was probably hampered by cycle-skipping artifacts for early arrivals recorded beyond 5 km of offset. However, the influence of the short-spread reflections and the redundancy is sufficiently high in the 3D data set, to guarantee the reconstruction of the main features of the target

4.4 Discussions and perspectives

We applied 3D acoustic FWI to the Valhall data set. Three-dimensional elastic FWI is the next challenge on Valhall. This application will require to account more accurately for anisotropy, density, and attenuation before reconstruction of the shear-wave velocity from the hydrophone and the geophone components (Prieux, 2012). As we saw, accounting for anisotropy is necessary to jointly match the kinematics of both short-spread reflections and diving waves in anisotropic environments. Moreover, in marine environment, the shear velocity can have a small influence in the data, and hence can be considered as a secondary parameter. It might be necessary to account for the influence of the other secondary parameters (density, attenuation) on the amplitudes of the data, to extract the information on the shear velocity from the data.

The main computational burden of FWI results from multi-source seismic modeling. Our FWI code was implemented with two levels of parallelism to reduce the computational time. This strategy requires however a significant number of processors. To reduce this demand, the simultaneous shooting techniques can be used. This approach builds a limited number of super shots by randomly encoding and stacking individual shots (Ben Hadj Ali *et al.*, 2011; Krebs *et al.*, 2009). A suitable stacking and encoding procedure, optimization algorithm, stopping criterion of iteration, and regularization should however be defined to make these approaches attractive. A second possible approach to reduce the number of modeling is to use the phase-sensitive detection approach proposed by Nihei et Li (2007). Modeling is performed in the time domain and frequency response is extracted by phase sensitive detection (basically, an extension of the discrete Fourier transform). If the source excitations are monochromatic functions, the frequency response of multiple sources can be extracted during one single time-domain modeling by encoding sources with slightly different frequencies.

FWI is an ill-posed and non-linear problem, that requires the starting model to be close enough to the real one. For future work, I shall test different initial models to assess the impact of the starting model. To mitigate the non-linearity of the problem, we shall consider

the hierarchical multiscale strategy. These strategies successively processes data subsets of increasing resolution power to incorporate smaller wavenumbers in the tomographic models by successive inversions of increasing frequencies. However, the starting frequency (or the maximum frequency of the group) must be sufficiently low to prevent cycle-skipping artifacts.

Our applications were performed without regularization. The regularizations can make the FWI better posed. We aim to implement multiplicative regularization in FWI (van den Berg *et al.*, 2003; Abubakar *et al.*, 2009) to automate as much as possible the weighting term, which controls the respective weight of the data-space and model-space misfit functions.

The $\mathcal{L}_2$ norm criterion used in our FWI algorithm shows some limitations, in instance with regards to noise. This norm approach assumes a Gaussian distribution of the misfit. We should perform careful quality control of the data to satisfy this assumption. Therefore, we will consider the other choice of the minimization criterion such as the absolute values norm $\mathcal{L}_1$, the Cauchy criterion norm and the hyperbolic secant criterion in FWI (Crase *et al.*, 1990; Pyun *et al.*, 2009; Brossier *et al.*, 2010c).

A series of mono-frequency inversion was performed from 3.5 Hz to 7 Hz using a frequency interval of 0.5 Hz. This implies that the first FWI model is built from the 3.5 Hz frequency only. Therefore, the FWI should be less sensitive to cycle-skipping artifacts, compared to the inversion of a [3.5 Hz - 4 Hz] frequency group (tests previously shown). The grid spacing is adapted to frequency and ranges from 70 m to 50 m. Note also that we perform 44 iterations instead of 27 for the previous tests presented in this study. We compare the FWI results obtained by mono-frequency inversions with that obtained by the frequency-group inversions (figures 4.1 and 4.2). We show similar results. However, the final FWI model inferred from the mono-frequency inversion is more contrasted. This might result, because the number of iteration was higher (44 instead of 27), and the volume of data injected during one inversion step is reduced, when mono-frequency inversion is considered. More importantly, the footprint of the suspicious low velocity zone at 1.2 km in depth is significantly reduced, and the agreement with the sonic log was nicely improved. The match of the early arrivals beyond 5 km of offset was also improved, and does not show evidence of cycle skipping artifacts anymore (Figure 4.3). These results show that the FWI set-up (here, the hierarchical management of frequencies in FWI) can have a significant influence on the FWI results, and further studies are required to establish a hierarchy between the factors that control the quality of the imaging. This is of particular importance in 3D FWI, because some of these factors, such as the number of frequencies, can have a significant impact on the computational cost of the 3D FWI.

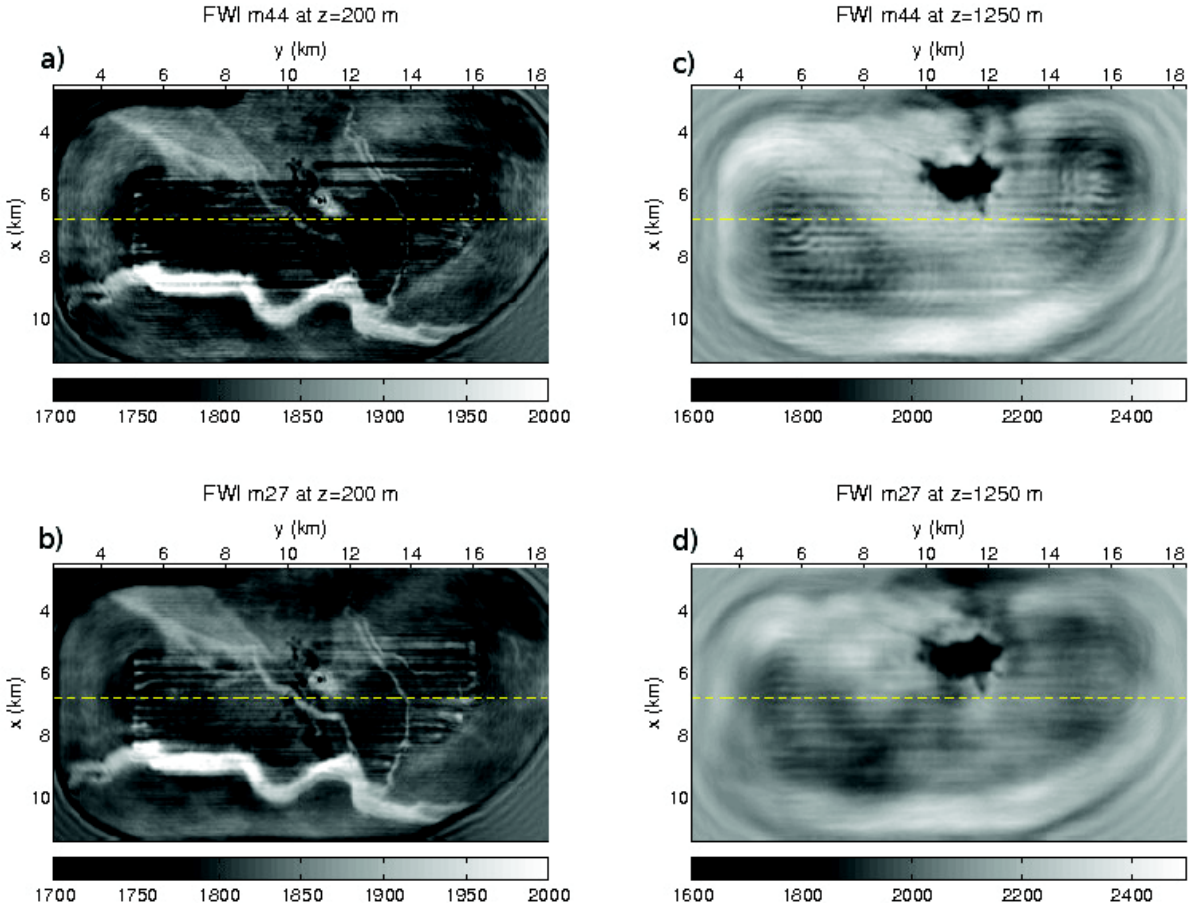

Figure 4.1: Comparison between the results of hierarchical mono-frequency and frequency-group inversions. (a, c) horizontal sections at 200 m and 1250 m in depth extracted from the final FWI model of the mono-frequency inversions. (b, d) Same as (a, c) for the final FWI model of the frequency-group inversion. Note that the footprint of the acquisition is more efficiently reduced in (b, d) compared to (a, c) when the number of frequencies involved in the inversion is increased and/or when multiple frequencies are simultaneously inverted.

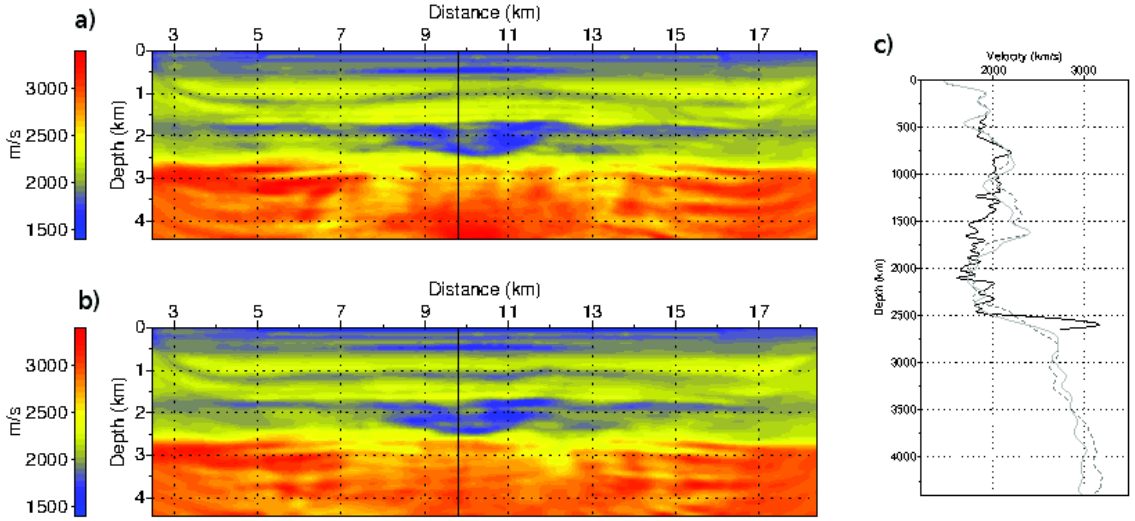

Figure 4.2: Comparison between the results of hierarchical mono-frequency and frequency-group inversions. (a, b) Vertical section at x = 6800 m extracted from the final FWI models of the mono-frequency (a) and frequency-group (b) inversions. (c) Vertical profile at x = 6800 m and y = 9500 m. Gray line corresponds to the frequency-group inversion results, dashed line corresponds to mono-frequency inversion, and black line is the sonic well log. Note how the footprint of the low-velocity layer at 1.2 km depth is reduced when sequential inversions of single frequencies are performed.

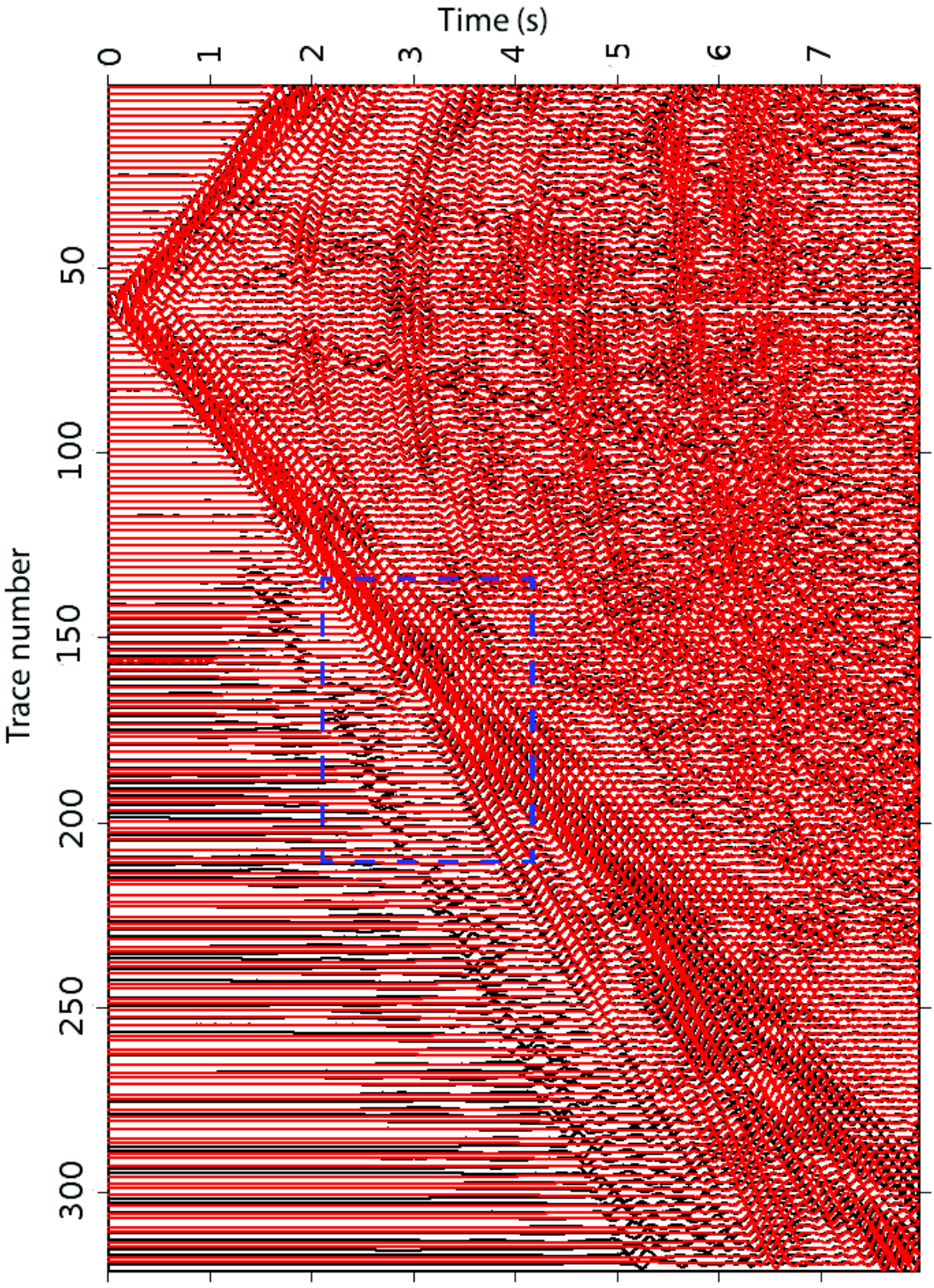

Figure 4.3: Direct comparison between recorded (black) seismograms and modeled (red) seismograms computed in the final FWI model of the hierarchical mono-frequency inversions. An automatic gain control with a window length of 2 s is applied to the seismograms, because we want to focus on the phase match. Note the good match of the first arrival beyond 5 km of offset.

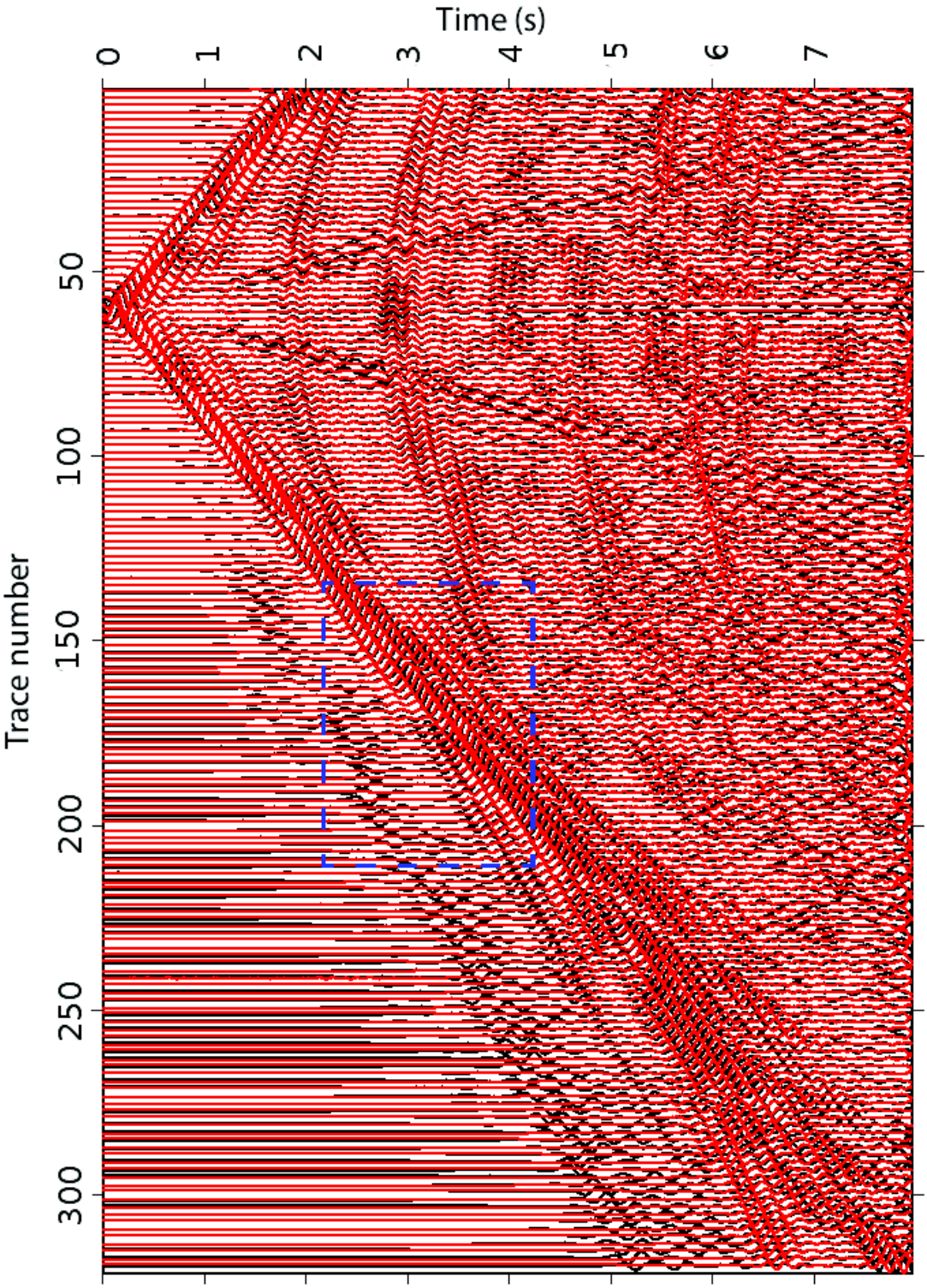

Figure 4.4: Same as figure 4.3, but the seismograms are computed in the final FWI model of the hierarchical frequency-group inversions. Note how the match of the first-arrival is degraded compared to the one shown in Figure 4.3. This confirms that FWI was hampered by cycle skipping artifacts during the first-frequency group inversion, which involves a maximum frequency of 4 Hz. These cycle skipping artifacts are reduced when the frequency of the first mono-frequency inversion is 3.5 Hz (Figure 4.3).

Bibliography

Abubakar, A., Hu, W., Habashy, T. M. et van den Berg, P. M. (2009). Application of the finite-difference contrast-source inversion algorithm to seismic full-waveform data. *Geophysics*, 74(6):WCC47–WCC58.

Aki, K. et Richards, P. (1980). *Quantitative Seismology: Theory and Methods*. W. H. Freeman & Co, San Francisco.

Alford, R., Kelly, K. et Boore, D. (1974a). Accuracy of finite-difference modeling of the acoustic wave equation. *Geophysics*, 39:834–842.

Alford, R. M., Kelly, K. R. et Boore, D. M. (1974b). Accuracy of finite-difference modeling of the acoustic wave equation. *Geophysics*, 39:834–842.

Alterman, Z. et Karal, F. C. (1968). Propagation of elastic waves in layared media by finite-difference methods. *Bulletin of the Seismological Society of America*, 58:367–398.

Amestoy, P., Davis, T. A. et Duff, I. S. (1996). An approximate minimum degree ordering algorithm. *SIAM Journal of Matrix Analysis and Applications*, 17:886–905.

Aminzadeh, F., Brac, J. et Kunz, T. (1997). *3-D Salt and Overthrust models*. SEG/EAGE 3-D Modeling Series No.1.

Aminzadeh, F., Burehard, N., Nicoletis, L., Rocca, F. et Wyatt, K. (1994). SEG/EAEG 3-D modeling project : Second report. *The Leading Edge*, 13:949–952.

Aminzadeh, F., Burehard, N., Nicoletis, L., Rocca, F. et Wyatt, K. (1995). SEG/EAEG 3-D modeling project : Third report. *The Leading Edge*, 14:125–128.

Ashcraft, C. et Liu, J. W. H. (1998). Robust ordering of sparse matrices using multisection. *SIAM Journal on Matrix Analysis and Applications*, 19(3):816–832.

Aubry, D. et Clouteau, D. (1991). A regularized boundary element method for stratified media. *In Proceedings of the First International Conference on Mathematical and Numerical Aspects of Wave Propagation Phenomena*, pages 660–668. SIAM, Philadephia.

Barkved, O., van Gestel, J., Bergsvik, L., Stockden, I. et Kommedal, J. (2009). Seismic plt - linking seismic time-lapse responses to production and injection data. *In Expanded Abstracts*.

Barnes, C. et Charara, M. (2009). The domain of applicability of acoustic full-waveform inversion for marine seismic data. *Geophysics*, 74(6):WCC91–WCC103.

Ben Hadj Ali, H. (2009). *Three dimensional visco-acoustic frequency domain full waveform inversion.* Thèse de doctorat, Université de Nice-Sophia-Antipolis.

Ben Hadj Ali, H., Operto, S. et Virieux, J. (2008). Velocity model building by 3D frequency-domain, full-waveform inversion of wide-aperture seismic data. *Geophysics*, 73(5):VE101–VE117.

Ben Hadj Ali, H., Operto, S. et Virieux, J. (2011). An efficient frequency-domain full waveform inversion method using simultaneous encoded sources. *Geophysics*, 76(4):R109.

Berenger, J.-P. (1994). A perfectly matched layer for absorption of electromagnetic waves. *Journal of Computational Physics*, 114:185–200.

Bohlen, T. et Saenger, E. H. (2006). Accuracy of heterogeneous staggered-grid finite-difference modeling of Rayleigh waves. *Geophysics*, 71:109–115.

Boore, D. M. (1972). Finite-difference methods for seismic wave propagation in heterogeneous materials. *In* Ed., B. B. A., éditeur : *Methods in computational physics*, volume 11. Academic Press, Inc.

Brenders, A. J. et Pratt, R. G. (2007). Full waveform tomography for lithospheric imaging: results from a blind test in a realistic crustal model. *Geophysical Journal International*, 168:133–151.

Brossier, R. (2009). *Imagerie sismique à deux dimensions des milieux visco-élastiques par inversion des formes d'onde: développements méthodologiques et applications.* Thèse de doctorat, Université de Nice-Sophia-Antipolis.

Brossier, R., Etienne, V., Operto, S. et Virieux, J. (2010a). Frequency-domain numerical modelling of visco-acoustic waves based on finite-difference and finite-element discontinuous galerkin methods. *In* Dissanayake, D. W., éditeur : *Acoustic Waves*, pages 125–158. SCIYO.

Brossier, R., Etienne, V., Operto, S. et Virieux, J. (2010b). Frequency-domain numerical modelling of visco-acoustic waves based on finite-difference and finite-element discontinuous galerkin methods. *In* Dissanayake, D. W., éditeur : *Acoustic Waves*, pages 125–158. SCIYO.

Brossier, R., Operto, S. et Virieux, J. (2009). Seismic imaging of complex onshore structures by 2D elastic frequency-domain full-waveform inversion. *Geophysics*, 74(6):WCC63–WCC76.

Brossier, R., Operto, S. et Virieux, J. (2010c). Which data residual norm for robust elastic frequency-domain full waveform inversion? *Geophysics*, 75(3):R37–R46.

Brugeas, I. (1996). Utilisation de MPI en décomposition de domaine. Rapport technique, IDRIS - CNRS. Publication IDRIS.

Bunks, C., Salek, F. M., Zaleski, S. et Chavent, G. (1995). Multiscale seismic waveform inversion. *Geophysics*, 60(5):1457–1473.

Byrd, R., Lu, P. et Nocedal, J. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific and Statistical Computing*, 16:1190–1208.

Castellanos, C., Etienne, V., Hu, G., Operto, S., Brossier, R. et Virieux, J. (2011). Algorithmic and methodological developments towards full waveform inversion in 3d elastic media. *SEG Technical Program Expanded Abstracts*, 30:2793–2798.

Cerjan, C., Kosloff, D., Kosloff, R. et Reshef, M. (1985). A nonreflecting boundary condition for discrete acoustic and elastic wave equations. *Geophysics*, 50(4):2117–2131.

Chavent, G. (2009). *Nonlinear least squares for inverse problems*. Springer Dordrecht Heidelberg London New York.

Clayton, R. et Engquist, B. (1977). Absorbing boundary conditions for acoustic and elastic wave equations. *Bulletin of the Seismological Society of America*, 67:1529–1540.

Collino, F. et Tsogka, C. (2001). Application of the perfectly matched absorbing layer model to the linear elastodynamic problem in anisotropic heterogeneous media. *Geophysics*, 66:294–307.

Courant, R., Friedrichs, K. et Lewy, H. (1928). Über die partiellen Differenzengleichungen der mathematischen Physik. *Mathematische Annalen*, 100:32–74.

Crase, E., Pica, A., Noble, M., McDonald, J. et Tarantola, A. (1990). Robust elastic non-linear waveform inversion: application to real data. *Geophysics*, 55:527–538.

Dablain, M. (1986). The application of high order differencing for the scalar wave equation. *Geophysics*, 51:54–66.

Danecek, P. et Seriani, G. (2008). An efficient parallel Chebyshev pseudo-spectral method for large-scale 3D seismic forward modelling. *In Expanded Abstracts*, page P046.

de la Puente, J., Dumbser, M., Käser, M. et Igel, H. (2008). Discontinuous Galerkin methods for wave propagation in poroelastic media. *Geophysics*, 73(5):77–97.

Dellinger, J. et Yu, J. (2009). Low-frequency virtual point-source interferometry using conventional sensors. *In Expanded Abstracts*.

Duff, I. S., Erisman, A. M. et Reid, J. K. (1986). *Direct methods for sparse matrices*. Clarendon Press, Oxford, U. K.

Duvaut, G. (1990). *Mécanique des milieux continus*. Masson.

Epanomeritakis, I., Akçelik, V., Ghattas, O. et Bielak, J. (2008). A Newton-CG method for large-scale three-dimensional elastic full waveform seismic inversion. *Inverse Problems*, 24:1–26.

Etienne, V., Hu, G., Operto, S., Virieux, J., Barkved, O. et Kommedal, J. (2012). Three-dimensional acoustic full waveform inversion: algorithm and application to Valhall. *In Expanded Abstracts, 74th Annual EAGE Conference & Exhibition, Copenhagen*. EAGE.

Etienne, V., Virieux, J., Hui, G., Jia, Y. et Operto, S. (2010). Computational and methodological developments towards 3D full waveform inversion. *In EOS Trans. AGU, abstract S31A-2049*. American Geophysical Union, San Francisco, USA.

Fichtner, A., Kennett, B. L. N., Igel, H. et Bunge, H. P. (2008). Theoretical background for continental- and global-scale full-waveform inversion in the time-frequency fomain. *Geophysical Journal International*, 175:665–685.

Fletcher, R. et Reeves, C. M. (1964). Function minimization by conjugate gradient. *Computer Journal*, 7:149–154.

Gauthier, O., Virieux, J. et Tarantola, A. (1986). Two-dimensional nonlinear inversion of seismic waveforms: numerical results. *Geophysics*, 51(7):1387–1403.

George, A. et Liu, J. W. (1981). *Computer solution of large sparse positive definite systems*. Prentice-Hall, Inc.

Gouveia, W. P. et Scales, J. A. (1998). Bayesian seismic waveform inversion: parameter estimation and uncertainty analysis. *Journal of Geophysical Research*, 103(B2):2579–2779.

Graves, R. (1996). Simulating seismic wave propagation in 3D elastic media using staggered-grid finite differences. *Bulletin of the Seismological Society of America*, 86:1091–1106.

Haidar, A. (2008). *On the parallel scalability of hybrid linear solvers for large 3D problems*. Thèse de doctorat, Institut National Polytechnique de Toulouse - CERFACS TH/PA/08/57.

Hansen, C. (1998). *Rank-defficient and discrete ill-posed problems - Numerical aspects of linear inversion*. Society for Industrial and Applied Mathematics - Mathematical modeling and computation.

Hanyga, A. et Pajchel, J. (1995). Point-to-curve ray tracing in complex geological models. *Geophysical Prospecting*, 43:859–872.

Hanyga, A. et Seredyńska, M. (1999). Asymptotic ray theory in poro- and viscoelastic media. *Wave Motion*, 30:175–195.

Hanyga, A., Thierry, P., Lambare, G. et Lucio, P. S. (1995). 2D and 3D asymptotic Green's functions for linear inversion. *In Proceedings*. SPIE.

Hicks, G. J. (2002). Arbitrary source and receiver positioning in finite-difference schemes using kaiser windowed sinc functions. *Geophysics*, 67:156–166.

Hustedt, B., Operto, S. et Virieux, J. (2004). Mixed-grid and staggered-grid finite difference methods for frequency domain acoustic wave modelling. *Geophysical Journal International*, 157:1269–1296.

Jaiswal, P., Zelt, C., Dasgupta, R. et Nath, K. (2009). Seismic imaging of the Naga Thrust using multiscale waveform inversion. *Geophysics*, 74(6):WCC129–WCC140.

Jo, C. H., Shin, C. et Suh, J. H. (1996). An optimal 9-point, finite-difference, frequency-space 2D scalar extrapolator. *Geophysics*, 61:529–537.

Kaiser, J. K. (1974). Nonrecursive digital filter design using the i0-sinh window function. *In Proceedings of the IEEE International Symposium on Circuits and Systems*, pages 20–23. IEEE.

Kelly, K., Ward, R., Treitel, S. et Alford, R. (1976). Synthetic seismograms - a finite-difference approach. *Geophysics*, 41:2–27.

Komatitsch, D. et Martin, R. (2007). An unsplit convolutional perfectly matched layer improved at grazing incidence for the seismic wave equation. *Geophysics*, 72(5):SM155–SM167.

Krebs, J., Anderson, J., Hinkley, D., Neelamani, R., Lee, S., Baumstein, A. et Lacasse, M. D. (2009). Fast full-wavefield seismic inversion using encoded sources. *Geophysics*, 74(6):WCC105–WCC116.

Krishnasamy, G., Rizzo, F. et Rudolphi, T. (1992). Hypersingular boundary integral equations: their occurrence, interpretation, regularization and computation. *In* Banerjee, P. et Kobayashi, S., éditeurs : *Developments in Boundary Element Methods*. Elsevier, New York.

Kuzuoglu, M. et Mittra, R. (1996). Frequency dependence of the constitutive parameters of causal perfectly matched anisotropic absorbers. *IEEE Microwave and Guided Wave Letters*, 6:447–449.

Lailly, P. (1984). The seismic inverse problem as a sequence of before stack migrations. *In* Bednar, R. et Weglein, éditeurs : *Conference on Inverse Scattering, SIAM, Philadelphia*, pages 206–220. Soc. Ind. appl. Math.

Levander, A. R. (1988). Fourth-order finite-difference P-SV seismograms. *Geophysics*, 53(11): 1425–1436.

Lions, J. (1972). *Nonhomogeneous boundary value problems and applications*. Springer Verlag, Berlin.

Malinowski, M., Operto, S. et Ribodetti, A. (2011). High-resolution seismic attenuation imaging from wide-aperture onshore data by visco-acoustic frequency-domain full waveform inversion. *Geophysical Journal International*, 186(3):1179–1204.

Marfurt, K. (1984). Accuracy of finite-difference anf finite-element modeling of the scalar and elastic wave equations. *Geophysics*, 49:533–549.

Menke, W. (1984). *Geophysical Data Analysis: Discrete Inverse Theory*. Academic Press, Inc., Orlando, USA.

Menke, W. et Abbott, D. (1990). *Geophysical Theory*. Columbia University Press, New York, USA.

Métivier, L., Brossier, R., Virieux, J. et Operto, S. (2012). Toward gauss-newton and exact newton optimization for full waveform inversion. *In EAGE, 74th Conference and Exhibition*, page This issue.

Miller, D., Oristaglio, M. et Beylkin, G. (1987). A new slant on seismic imaging: Migration and integral geometry. *Geophysics*, 52(7):943–964.

Min, D.-J., Shin, C., Pratt, R. G. et Yoo, H. S. (2003). Weighted-averaging finite-element method for 2D elastic wave equations in the frequency domain. *Bulletin of the Seismological Society of America*, 93(2):904–921.

Moczo, P., Kristek, J. et Halada, L. (2000). 3D fourth-order staggered-grid finite-difference schemes: stability and dispersion. *Bulletin of the Seismological Society of America*, 90:587–603.

Mora, P. R. (1987). Nonlinear two-dimensional elastic inversion of multi-offset seismic data. *Geophysics*, 52:1211–1228.

Mulder, W. et Plessix, R. E. (2008). Exploring some issues in acoustic full waveform inversion. *Geophysical Prospecting*, 56(6):827–841.

Munns, J. W. (1985). The Valhall field: a geological overview. *Marine and Petroleum Geology*, 2:23–43.

Nihei, K. T. et Li, X. (2007). Frequency response modelling of seismic waves using finite difference time domain with phase sensitive detection (TD-PSD). *Geophysical Journal International*, 169:1069–1078.

Nocedal, J. (1980). Updating Quasi-Newton Matrices With Limited Storage. *Mathematics of Computation*, 35(151):773–782.

Nocedal, J. et Wright, S. J. (1999). *Numerical Optimization*. New York, US : Springer.

O. Barkved, P. H. (2003). Society of petroleum engineers inc. valhall field - still on plateau after 20 years of production. *This paper was prepared for presentation at Offshore Europe 2003 held in Aberdeen, UK.*

Olofsson, B., Probert, T., Kommedal, J. et Barkved, O. (2003). Azimuthal anisotropy from the valhall 4C 3D survey. *The Leading Edge*, pages 1228–1235.

Operto, S. (2006). 3D frequency-domain finite-difference modeling of acoustic wave propagation using the massively parallel solver MUMPS. Rapport technique, UMR Géosciences Azur.

Operto, S., Virieux, J., Amestoy, P., L'Éxcellent, J.-Y., Giraud, L. et Ben Hadj Ali, H. (2007). 3D finite-difference frequency-domain modeling of visco-acoustic wave propagation using a massively parallel direct solver: A feasibility study. *Geophysics*, 72(5):SM195–SM211.

Operto, S., Virieux, J., Dessa, J. X. et Pascal, G. (2006). Crustal imaging from multifold ocean bottom seismometers data by frequency-domain full-waveform tomography: application to the eastern Nankai trough. *Journal of Geophysical Research*, 111(B09306): doi:10.1029/2005JB003835.

Operto, S., Virieux, J., Hustedt, B. et Malfanti, F. (2002). Adaptative wavelet-based finite-difference modelling of SH-wave propagation. *Geophysical Journal International*, 148: 1–28.

Operto, S., Virieux, J., Ribodetti, A. et Anderson, J. E. (2009). Finite-difference frequency-domain modeling of visco-acoustic wave propagation in two-dimensional TTI media. *Geophysics*, 74 (5):T75–T95.

Plessix, R. E. (2006). A review of the adjoint-state method for computing the gradient of a functional with geophysical applications. *Geophysical Journal International*, 167(2):495–503.

Plessix, R. E. (2007). A Helmholtz iterative solver for 3D seismic-imaging problems. *Geophysics*, 72(5):SM185–SM194.

Plessix, R. E. (2009). Three-dimensional frequency-domain full-waveform inversion with an iterative solver. *Geophysics*, 74(6):WCC53–WCC61.

Plessix, R. E. (2012). Waveform inversion overview: Where are we? and what are the challeges? *In Expanded Abstracts*. EAGE.

Plessix, R. E. et Perkins, C. (2010). Full waveform inversion of a deep water ocean bottom seismometer dataset. *First Break*, 28:71–78.

Podvin, P. et Lecomte, I. (1991). Finite difference computation of traveltimes in very contrasted velocity model : a massively parallel approach and its associated tools. *Geophysical Journal International*, 105:271–284.

Polak, E. et Ribière, G. (1969). Note sur la convergence de méthodes de directions conjuguées. *Revue Française d'Informatique et de Recherche Opérationnelle*, 16:35–43.

Pratt, R. G. (1990a). Frequency-domain elastic modeling by finite differences: a tool for crosshole seismic imaging. *Geophysics*, 55(5):626–632.

Pratt, R. G. (1990b). Inverse theory applied to multi-source cross-hole tomography. part II : elastic wave-equation method. *Geophysical Prospecting*, 38:311–330.

Pratt, R. G. (1999). Seismic waveform inversion in the frequency domain, part I : theory and verification in a physic scale model. *Geophysics*, 64:888–901.

Pratt, R. G. (2004). Velocity models from frequency-domain waveform tomography: past, present and future. *In Expanded Abstracts*. Eur. Ass. Expl. Geophys.

Pratt, R. G. (2008). Waveform tomography - successes, cautionary tales, and future directions. *In Presented at the 70th Annual EAGE Conference & Exhibition, Roma*, pages WO11 – Full–Waveform Inversion: current status and perspectives.

Pratt, R. G. et Sams, M. S. (1996). Reconciliation of crosshole seismic velocities with well information in a layered sedimentary environment. *Geophysics*, 61:549–560.

Pratt, R. G., Shin, C. et Hicks, G. J. (1998). Gauss-Newton and full Newton methods in frequency-space seismic waveform inversion. *Geophysical Journal International*, 133:341–362.

Pratt, R. G. et Worthington, M. H. (1990). Inverse theory applied to multi-source cross-hole tomography. Part I: acoustic wave-equation method. *Geophysical Prospecting*, 38:287–310.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. et Flannery, B. P. (1992). *Numerical recipes in FORTRAN : the art of scientific computing; second edition*. Cambridge university press.

Prieux, V. (2012). *Imagerie sismique des milieux visco-acoustiques et visco-élastiques à deux dimensions par stéréotomographie et inversion des formes d'ondes: applications au champ pétrolier de Valhall*. Thèse de doctorat, Université de Nice-Sophia-Antipolis.

Prieux, V., Brossier, R., Gholami, Y., Operto, S., Virieux, J., Barkved, O. et Kommedal, J. (2011). On the footprint of anisotropy on isotropic full waveform inversion: the Valhall case study. *Geophysical Journal International*, 187:1495–1515.

Prieux, V., Operto, S., Brossier, R. et Virieux, J. (2009). Application of acoustic full waveform inversion to the synthetic Valhall model. volume 28, pages 2268–2272. SEG.

Pyun, S., Shin, C. et Son, W. (2009). Frequency-domain waveform inversion using an L1-norm objective function. *In Expanded Abstracts*, page P005. EAGE.

Ravaut, C., Operto, S., Improta, L., Virieux, J., Herrero, A. et dell'Aversana, P. (2004). Multi-scale imaging of complex structures from multi-fold wide-aperture seismic data by frequency-domain full-wavefield inversions: application to a thrust belt. *Geophysical Journal International*, 159:1032–1056.

Riyanti, C. D., Kononov, A., Erlangga, Y. A., Vuik, C., Oosterlee, C., Plessix, R. E. et Mulder, W. A. (2007). A parallel multigrid-based preconditioner for the 3D heterogeneous high-frequency Helmholtz equation. *Journal of Computational physics*, 224: 431–448.

Royer, D. et Dieulesaint, E. (1997). *Ondes élastiques dans les solides*. Masson.

Saenger, E. H., Gold, N. et Shapiro, S. A. (2000). Modeling the propagation of elastic waves using a modified finite-difference grid. *Wave motion*, 31:77–92.

Sambridge, M. S., Tarantola, A. et Kennett, B. L. (1991). An alternative strategy for non-linear inversion of seismic waveforms. *Geophysical Prospecting*, 39:723–736.

Scales, J. A., Docherty, P. et Gersztenkorn, A. (1990). Regularization of nonlinear inverse problems: imaging the near-surface weathering layer. *Inverse Problems*, 6:115–131.

Scales, J. A. et Smith, M. L. (1994). *Introductory geophysical inverse theory*. Samizdat press.

Sears, T., Singh, S. et Barton, P. (2008). Elastic full waveform inversion of multi-component OBC seismic data. *Geophysical Prospecting*, 56(6):843–862.

Seriani, G. et Priolo, E. (1994). Spectral element method for acoustic wave simulation in heterogeneous media. *Finite elements in analysis and design*, 16:337–348.

Shearer, P. M. (2009). *Introduction to seismology*. Cambridge University Press, Cambridge, UK.

Sheng, J., Leeds, A., Buddensiek, M. et Schuster, G. T. (2006). Early arrival waveform tomography on near-surface refraction data. *Geophysics*, 71(4):U47–U57.

Shi, Y., Zhao, W. et Cao, H. (2007). Nonlinear process control of wave-equation inversion and its application in the detection of gas. *Geophysics*, 72(1):R9–R18.

Shin, C. et Cha, Y. H. (2008). Waveform inversion in the Laplace domain. *Geophysical Journal International*, 173(3):922–931.

Shin, C., Jang, S. et Min, D. J. (2001a). Improved amplitude preservation for prestack depth migration by inverse scattering theory. *Geophysical Prospecting*, 49:592–606.

Shin, C., Min, D.-J., Marfurt, K. J., Lim, H. Y., Yang, D., Cha, Y., Ko, S., Yoon, K., Ha, T. et Hong, S. (2002). Traveltime and amplitude calculations using the damped wave solution. *Geophysics*, 67:1637–1647.

Shin, C., Yoon, K., Marfurt, K. J., Park, K., Yang, D., Lim, H. Y., Chung, S. et Shin, S. (2001b). Efficient calculation of a partial derivative wavefield using reciprocity for seismic imaging and inversion. *Geophysics*, 66(6):1856–1863.

Sirgue, L. (2006). The importance of low frequency and large offset in waveform inversion. *In Presented at the 68th EAGE Conference & Exhibition, Vienna, EAGE*, page A037.

Sirgue, L., Barkved, O. I., Dellinger, J., Etgen, J., Albertin, U. et Kommedal, J. H. (2010). Full waveform inversion: the next leap forward in imaging at Valhall. *First Break*, 28:65–70.

Sirgue, L., Barkved, O. I., Gestel, J. P. V., Askim, O. J. et Kommedal, J. H. . (2009). 3D waveform inversion on Valhall wide-azimuth OBC. *In Presented at the 71th Annual International Meeting, EAGE, Expanded Abstracts*, page U038.

Sirgue, L., Etgen, J. T. et Albertin, U. (2008). 3D Frequency Domain Waveform Inversion using Time Domain Finite Difference Methods. *In Proceedings 70th EAGE, Conference and Exhibition, Roma, Italy*, page F022.

Sirgue, L. et Pratt, R. G. (2004). Efficient waveform inversion and imaging : a strategy for selecting temporal frequencies. *Geophysics*, 69(1):231–248.

Soubaras, R. et Whiting, P. (2011). Variable depth streamer — the new broadband acquisition system. *SEG Technical Program Expanded Abstracts*, 30(1):4349–4353.

Sourbier, F., Haiddar, A., Giraud, L., Ben-Hadj-Ali, H., Operto, S. et Virieux, J. (2011). Three-dimensional parallel frequency-domain visco-acoustic wave modelling based on a hybrid direct/iterative solver. *Geophysical Prospecting*, 59(5):834–856.

Sourbier, F., Haiddar, A., Giraud, L., Brossier, R., Operto, S. et Virieux, J. (2008). Frequency-domain full-waveform modeling using a hybrid direct-iterative solver based on a parallel domain decomposition method. *In 70th Annual EAGE Conference & Exhibition, Roma*. Eur. Ass. Expl.Geophys.

Sourbier, F., Operto, S., Virieux, J., Amestoy, P. et L'Excellent, J.-Y. (2009a). FWT2D: A massively parallel program for frequency-domain full-waveform tomography of wide-aperture seismic data–Part 1: Algorithm. *Computers & Geosciences*, 35(3):487 – 495.

Sourbier, F., Operto, S., Virieux, J., Amestoy, P. et L'Excellent, J.-Y. (2009b). FWT2D: A massively parallel program for frequency-domain full-waveform tomography of wide-aperture seismic data–Part 2: Numerical examples and scalability analysis. *Computers & Geosciences*, 35(3):496 – 514.

Stekl, I. et Pratt, R. G. (1998). Accurate viscoelastic modeling by frequency-domain finite difference using rotated operators. *Geophysics*, 63:1779–1794.

Tarantola, A. (1984a). Inversion of seismic reflection data in the acoustic approximation. *Geophysics*, 49(8):1259–1266.

Tarantola, A. (1984b). Linearized inversion of seismic reflection data. *Geophysical Prospecting*, 32:998–1015.

Tarantola, A. (1987). *Inverse problem theory: methods for data fitting and model parameter estimation*. Elsevier, New York.

Thierry, P., Operto, S. et Lambaré, G. (1999). Fast 2D ray-Born inversion/migration in complex media. *Geophysics*, 64(1):162–181.

Tikhonov, A. et Arsenin, V. (1977). *Solution of ill-posed problems*. Winston, Washington, DC.

Tikhonov, A. N. (1963). Resolution of ill-posed problems and the regularization method (in russian, french translation, mir, moscow, 1976). *Dokl. Akad. Nauk SSSR*, 151:501–504.

van den Berg, P., Abubakar, A. et Fokkema, J. (2003). Multiplicative regularization for contrast profile inversion. *Radio Science*, 38:23.1–23.10.

van Gestel, J., Best, K., Barkved, O. et Kommedal, J. (2008). Integrating frequent time-lapse data into the reservoir simulation modeling of the valhall field. *In Expanded Abstracts*.

Virieux, J. (1984). SH wave propagation in heterogeneous media, velocity-stress finite difference method. *Geophysics*, 49:1259–1266.

Virieux, J. (1986). P-SV wave propagation in heterogeneous media, velocity-stress finite difference method. *Geophysics*, 51:889–901.

Virieux, J. et Madariaga, R. (1982). Dynamic faulting studied by a finite difference method. *Bulletin of the Seismological Society of America*, 72:345–369.

Virieux, J. et Operto, S. (2009). An overview of full waveform inversion in exploration geophysics. *Geophysics*, 74(6):WCC127–WCC152.

Virieux, J., Operto, S., Ben Hadj Ali, H., Brossier, R., Etienne, V., Sourbier, F., Giraud, L. et Haidar, A. (2009). Seismic wave modeling for seismic imaging. *The Leading Edge*, 28(5):538–544.

Wang, S., de Hoop, M. V. et Xia, J. (2010). Seismic inverse scattering via Helmholtz operator factorization and optimization. *Journal of Computational Physics*, 229:8445–8462.

Wang, S., de Hoop, M. V. et Xia, J. (2011a). On 3d modeling of seismic wave propagation via a structured parallel multifrontal direct helmholtz solver. *Geophysical Prospecting*, 59(5):857–873.

Wang, S., Xia, J., Situ, Y., de Hoop, M. V. et Li, X. (2011b). Efficient parallel algorithms for hierarchically semiseparable (hss) matrices: Kernel of a massively parallel structured direct helmholtz solver. *SEG Technical Program Expanded Abstracts*, 30(1):2523–2527.

Wu, R. S. et Toksöz, M. N. (1987). Diffraction tomography and multisource holography applied to seismic imaging. *Geophysics*, 52:11–25.

Yee, K. S. (1966). Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media. *IEEE Transactions on Antennas and Propagation*, 14:302–307.

Zhang, Y.-G. et Ballmann, J. (1997). Two techniques for the absorption of elastic waves using an artificial transition layer. *Wave motion*, 25:15–33.